Random sampling locations for comparing a mean with a fixed threshold (parametric)

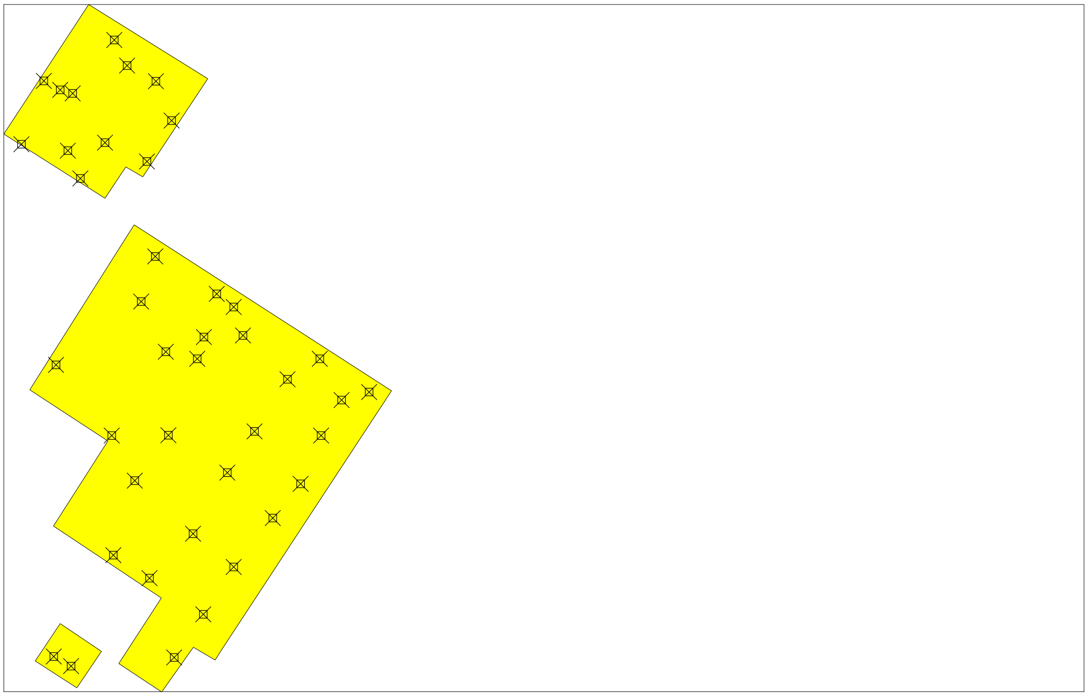
Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	11
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$6,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.

9557822



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	1110	Manual	T
679301.1600	3083254.0340	J-12S	748	Manual	T
679171.2970	3083289.7960	J-04S	747	Manual	T
679155.0740	3083294.6960	J-03S	2250	Manual	T
679133.4290	3083306.3130	J-01S	1200	Manual	T
679104.2450	3083223.2620	J-02S	7690	Manual	T
679164.8060	3083214.7100	J-06S	6880	Manual	T
679181.2750	3083178.2880	J-08S	10900	Manual	T
679213.7730	3083224.9730	J-09S	858	Manual	T
679280.5440	3083305.6810	J-10S	1090	Manual	T
679242.7260	3083326.5280	J-07S	11400	Manual	T
679225.8560	3083359.9740	J-05S	959	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	5880	Manual	T
679335.0020	3082941.1720	J-21S	648	Manual	T
679343.5810	3082969.5980	J-19S	1760	Manual	T
679394.8070	3082971.8300	J-24S	790	Manual	T
679382.8640	3083009.1130	J-20S	915	Manual	T
679293.5600	3082950.4980	J-17S	1130	Manual	T
679360.5700	3083026.4980	J-18S	3510	Manual	T

679297.0010	3082840.6970	J-23S	4180	Manual	T
679252.7130	3082781.0290	J-22S	1960	Manual	T
679222.6340	3082840.1720	J-16S	3640	Manual	T
679279.6830	3083075.4290	J-14S	917	Manual	T
679149.4920	3082933.0980	J-13S	5140	Manual	T
679272.0040	3082652.6750	J-28S	9150	Manual	T
679224.5850	3082683.1400	J-26S	2800	Manual	T
679146.6460	3082549.7640	J-25S	13800	Manual	T
679169.0760	3082537.3510	J-27S	4660	Manual	T
679342.7410	3082605.3190	J-35S	2140	Manual	T
679304.6530	3082548.6880	J-34S	1830	Manual	T
679382.8900	3082667.5270	J-36S	7640	Manual	T
679329.4380	3082711.0960	J-29S	3630	Manual	T
679374.4420	3082791.3300	J-30S	1770	Manual	T
679453.4760	3082914.1150	J-32S	2460	Manual	T
679410.1490	3082845.8460	J-31S	1390	Manual	T
679560.6070	3082897.2580	J-41S	2480	Manual	T
679495.8840	3082940.9730	J-33S	1240	Manual	T
679524.3310	3082886.8990	J-40S	1495	Manual	T
679497.3310	3082840.3960	J-39S	1160	Manual	T
679470.3570	3082776.7350	J-38S	3310	Manual	T
679433.9450	3082731.6820	J-37S	8450	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

The values of these inputs that result in the calculated number of sampling locations are:

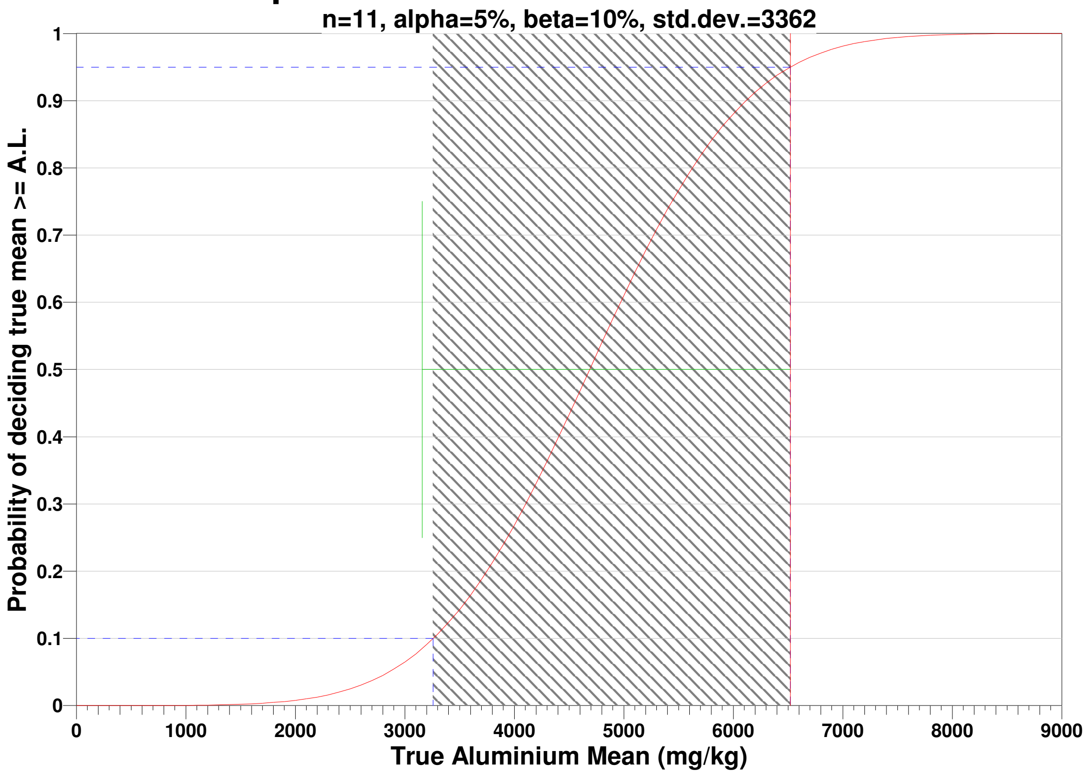
Analyte	**n**	**Parameter**					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Aluminium	11	3362 mg/kg	3261 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.

1-Sample t-Test of True Mean vs. Action Level



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=6521		α=5		α=10		α=15	
		s=6724	s=3362	s=6724	s=3362	s=6724	s=3362
LBGR=90	β=5	1152	290	912	229	765	192
	β=10	912	229	700	176	572	144
	β=15	766	193	573	144	458	115
LBGR=80	β=5	290	74	229	58	192	49
	β=10	229	59	176	45	144	37
	β=15	193	50	144	37	115	30
LBGR=70	β=5	130	34	102	27	86	22

	β=10	103	27	79	21	65	17
	β=15	87	23	65	17	52	14

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $6,500.00, which averages out to a per sample cost of $590.91. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**11 Samples**
Field collection costs		$100.00	$1,100.00
Analytical costs	$400.00	$400.00	$4,400.00
Sum of Field & Analytical costs		**$500.00**	**$5,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$6,500.00**

Data Analysis for Aluminium
The following data points were entered by the user for analysis.

Aluminium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	648	747	748	790	858	915	917	959	1090	1110
10	1130	1160	1200	1240	1390	1495	1760	1770	1830	1960
20	2140	2250	2460	2480	2800	3310	3510	3630	3640	4180
30	4660	5140	5880	6880	7640	7690	8450	9150	1.09e+004	1.14e+004
40	1.38e+004									

SUMMARY STATISTICS for Aluminium	
n	41
Min	648
Max	13800
Range	13152
Mean	3553.8
Median	2140
Variance	1.1301e+007
StdDev	3361.7
Std Error	525
Skewness	1.4858

Interquartile Range				3780				
Percentiles								
1%	5%	10%	25%	50%	75%	90%	95%	99%
648	747.1	803.6	1120	2140	4900	9010	1.135e+004	1.38e+004

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Aluminium			
k	Test Statistic R_k	5% Critical Value C_k	Significant?
1	3.048	3.05	No

None of the test statistics exceeded the corresponding critical values, therefore none of the 1 tests are significant and we conclude that at the 5% significance level there are no outliers in the data.

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.8
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

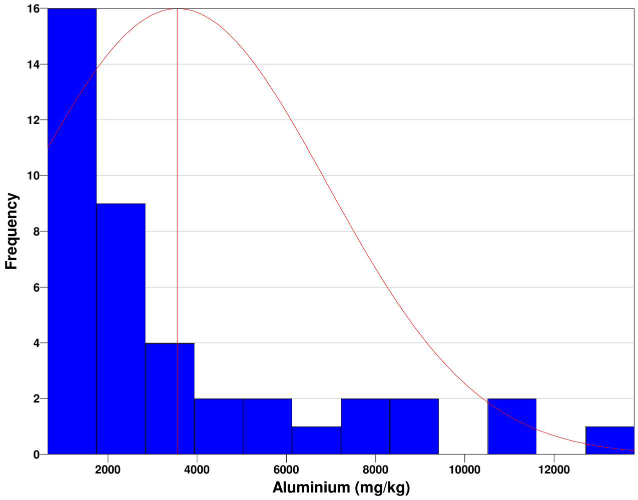
Data Plots for Aluminium
Graphical displays of the data are shown below.

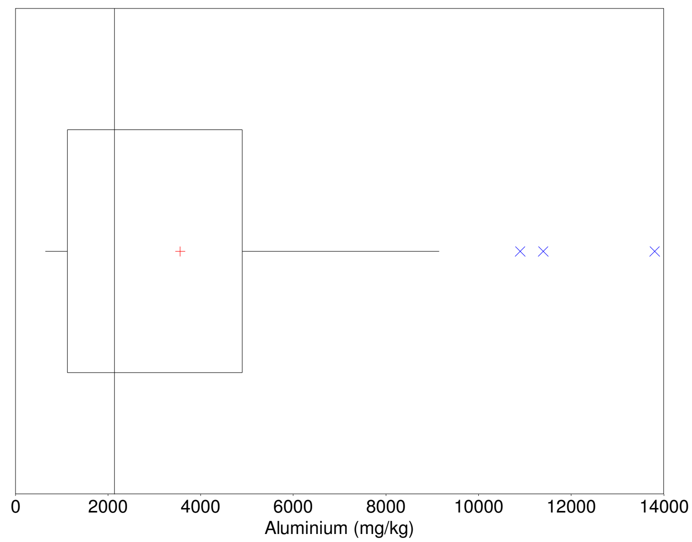
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

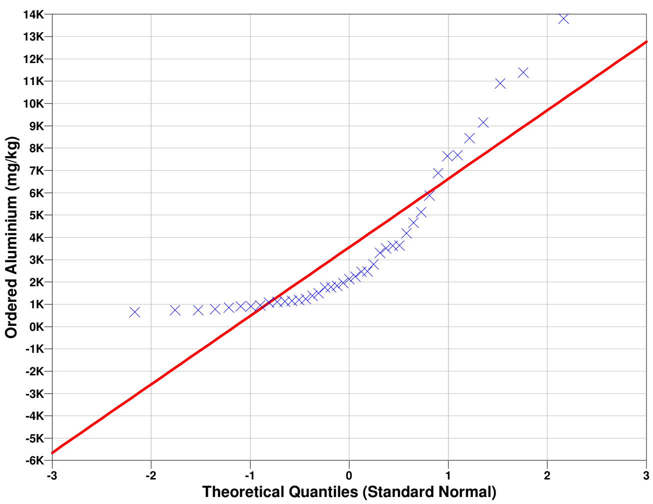
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the

distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Aluminium

A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.7943
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	4438
95% Non-Parametric (Chebyshev) UCL	5842

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (5842) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (6521),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-5.6517	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
33	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

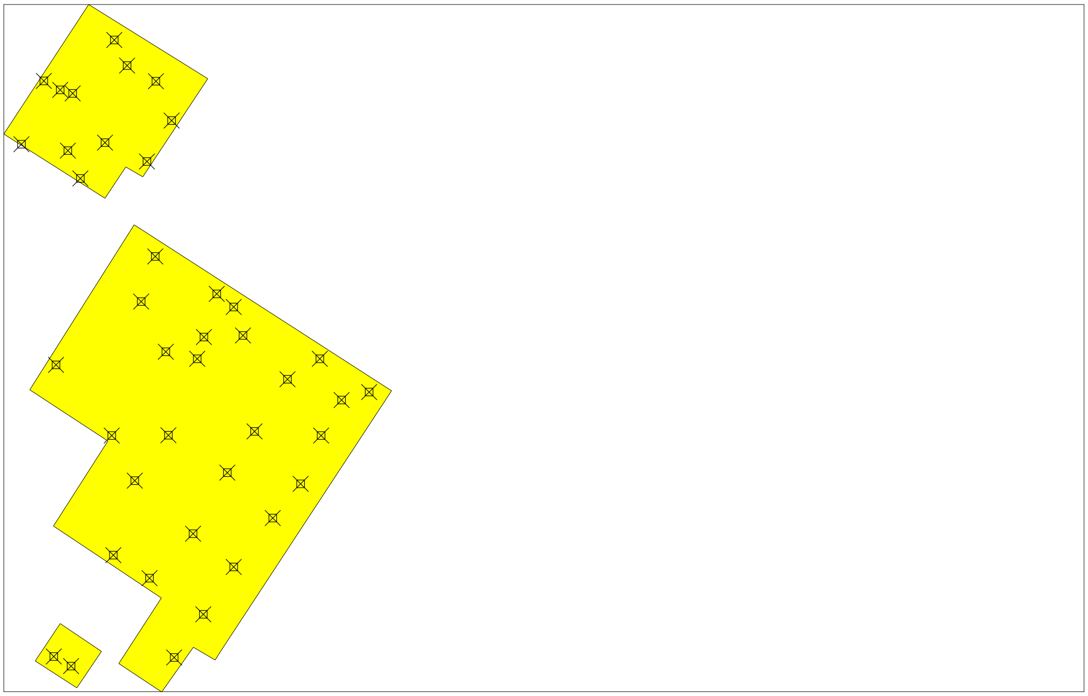
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	13
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$7,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	1110	Manual	T
679301.1600	3083254.0340	J-12S	748	Manual	T
679171.2970	3083289.7960	J-04S	747	Manual	T
679155.0740	3083294.6960	J-03S	2250	Manual	T
679133.4290	3083306.3130	J-01S	1200	Manual	T
679104.2450	3083223.2620	J-02S	7690	Manual	T
679164.8060	3083214.7100	J-06S	6880	Manual	T
679181.2750	3083178.2880	J-08S	10900	Manual	T
679213.7730	3083224.9730	J-09S	858	Manual	T
679280.5440	3083305.6810	J-10S	1090	Manual	T
679242.7260	3083326.5280	J-07S	11400	Manual	T
679225.8560	3083359.9740	J-05S	959	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	5880	Manual	T
679335.0020	3082941.1720	J-21S	648	Manual	T
679343.5810	3082969.5980	J-19S	1760	Manual	T
679394.8070	3082971.8300	J-24S	790	Manual	T
679382.8640	3083009.1130	J-20S	915	Manual	T
679293.5600	3082950.4980	J-17S	1130	Manual	T
679360.5700	3083026.4980	J-18S	3510	Manual	T

679297.0010	3082840.6970	J-23S	4180	Manual	T
679252.7130	3082781.0290	J-22S	1960	Manual	T
679222.6340	3082840.1720	J-16S	3640	Manual	T
679279.6830	3083075.4290	J-14S	917	Manual	T
679149.4920	3082933.0980	J-13S	5140	Manual	T
679272.0040	3082652.6750	J-28S	9150	Manual	T
679224.5850	3082683.1400	J-26S	2800	Manual	T
679146.6460	3082549.7640	J-25S	13800	Manual	T
679169.0760	3082537.3510	J-27S	4660	Manual	T
679342.7410	3082605.3190	J-35S	2140	Manual	T
679304.6530	3082548.6880	J-34S	1830	Manual	T
679382.8900	3082667.5270	J-36S	7640	Manual	T
679329.4380	3082711.0960	J-29S	3630	Manual	T
679374.4420	3082791.3300	J-30S	1770	Manual	T
679453.4760	3082914.1150	J-32S	2460	Manual	T
679410.1490	3082845.8460	J-31S	1390	Manual	T
679560.6070	3082897.2580	J-41S	2480	Manual	T
679495.8840	3082940.9730	J-33S	1240	Manual	T
679524.3310	3082886.8990	J-40S	1495	Manual	T
679497.3310	3082840.3960	J-39S	1160	Manual	T
679470.3570	3082776.7350	J-38S	3310	Manual	T
679433.9450	3082731.6820	J-37S	8450	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

The values of these inputs that result in the calculated number of sampling locations are:

Analyte	**n**	**Parameter**					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Aluminium	13	3362 mg/kg	2967 mg/kg	0.05	0.1	1.64485	1.28155

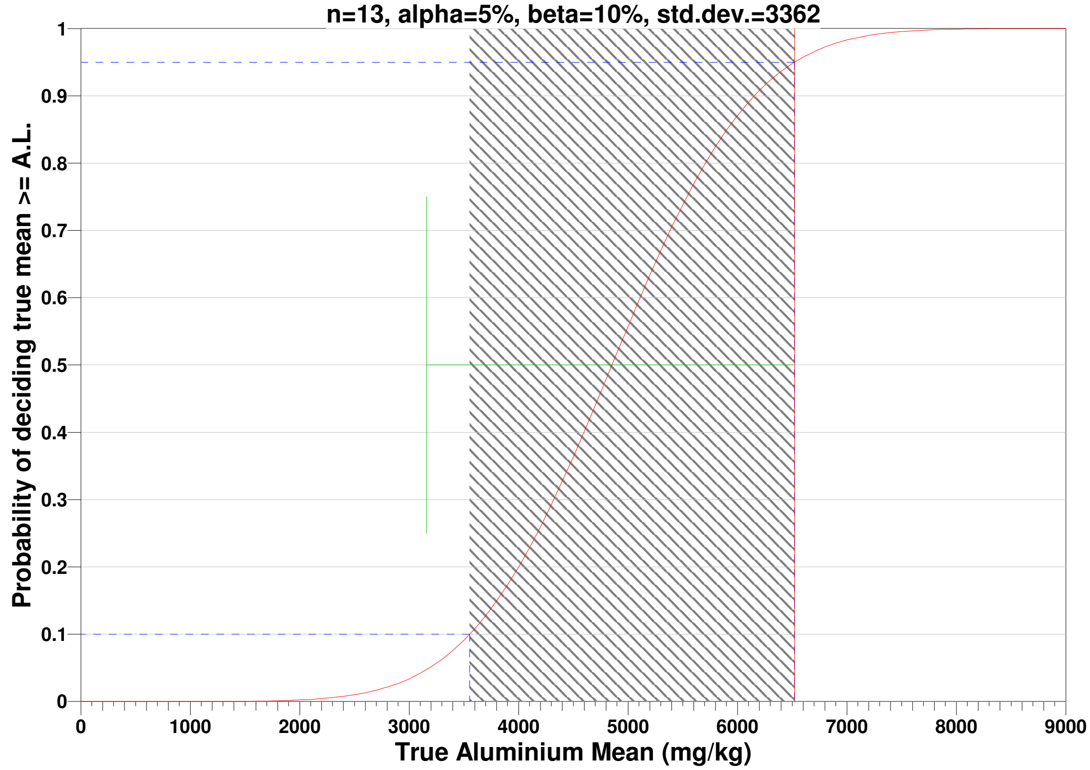
[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.

1-Sample t-Test of True Mean vs. Action Level

n=13, alpha=5%, beta=10%, std.dev.=3362



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=6521		α=5		α=10		α=15	
		s=6724	s=3362	s=6724	s=3362	s=6724	s=3362
LBGR=90	β=5	1152	290	912	229	765	192
	β=10	912	229	700	176	572	144
	β=15	766	193	573	144	458	115
LBGR=80	β=5	290	74	229	58	192	49
	β=10	229	59	176	45	144	37
	β=15	193	50	144	37	115	30
LBGR=70	β=5	130	34	102	27	86	22

	β=10	103	27	79	21	65	17
	β=15	87	23	65	17	52	14

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $7,500.00, which averages out to a per sample cost of $576.92. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**13 Samples**
Field collection costs		$100.00	$1,300.00
Analytical costs	$400.00	$400.00	$5,200.00
Sum of Field & Analytical costs		**$500.00**	**$6,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$7,500.00**

Data Analysis for Aluminium
The following data points were entered by the user for analysis.

Aluminium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	648	747	748	790	858	915	917	959	1090	1110
10	1130	1160	1200	1240	1390	1495	1760	1770	1830	1960
20	2140	2250	2460	2480	2800	3310	3510	3630	3640	4180
30	4660	5140	5880	6880	7640	7690	8450	9150	1.09e+004	1.14e+004
40	1.38e+004									

SUMMARY STATISTICS for Aluminium	
n	41
Min	648
Max	13800
Range	13152
Mean	3553.8
Median	2140
Variance	1.1301e+007
StdDev	3361.7
Std Error	525
Skewness	1.4858

Interquartile Range				3780				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
648	747.1	803.6	1120	2140	4900	9010	1.135e+004	1.38e+004

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Aluminium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.048	3.05	No

None of the test statistics exceeded the corresponding critical values, therefore none of the 1 tests are significant and we conclude that at the 5% significance level there are no outliers in the data.

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.8
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

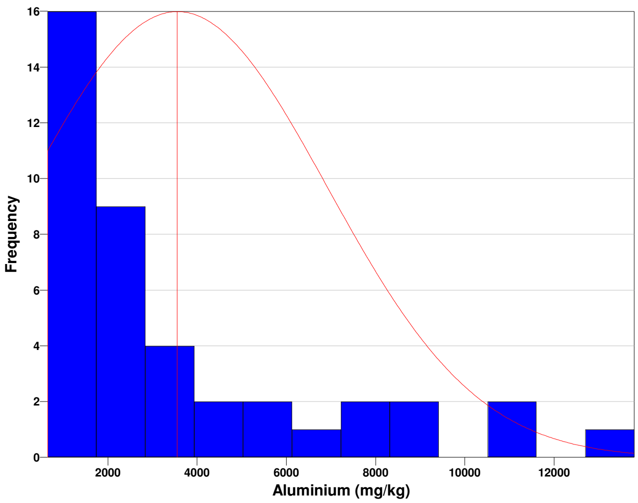
Data Plots for Aluminium
Graphical displays of the data are shown below.

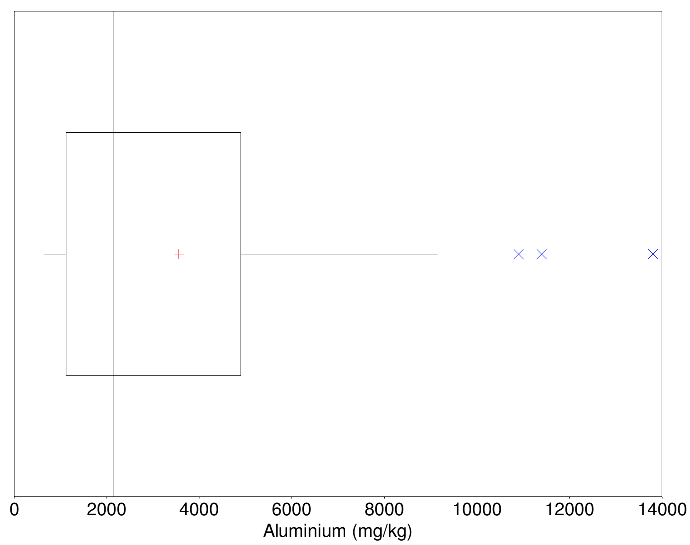
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

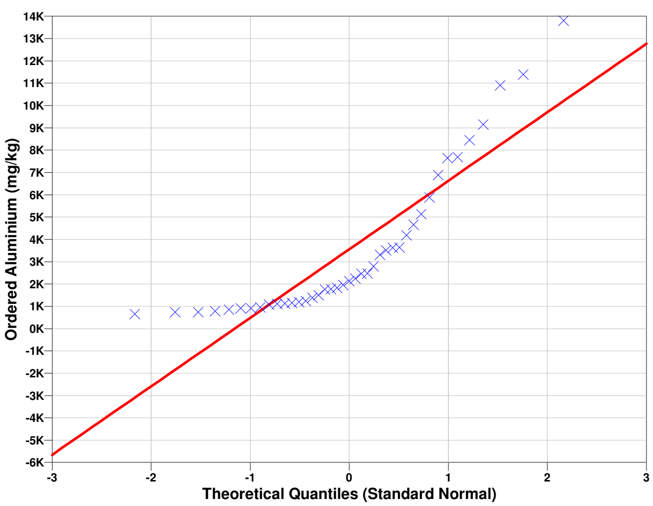
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the

distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Aluminium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.7943
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	4438
95% Non-Parametric (Chebyshev) UCL	5842

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (5842) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (6521),
SE is the standard error = (standard deviation) / (square root of n).

This *t* was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-5.6517	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
33	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

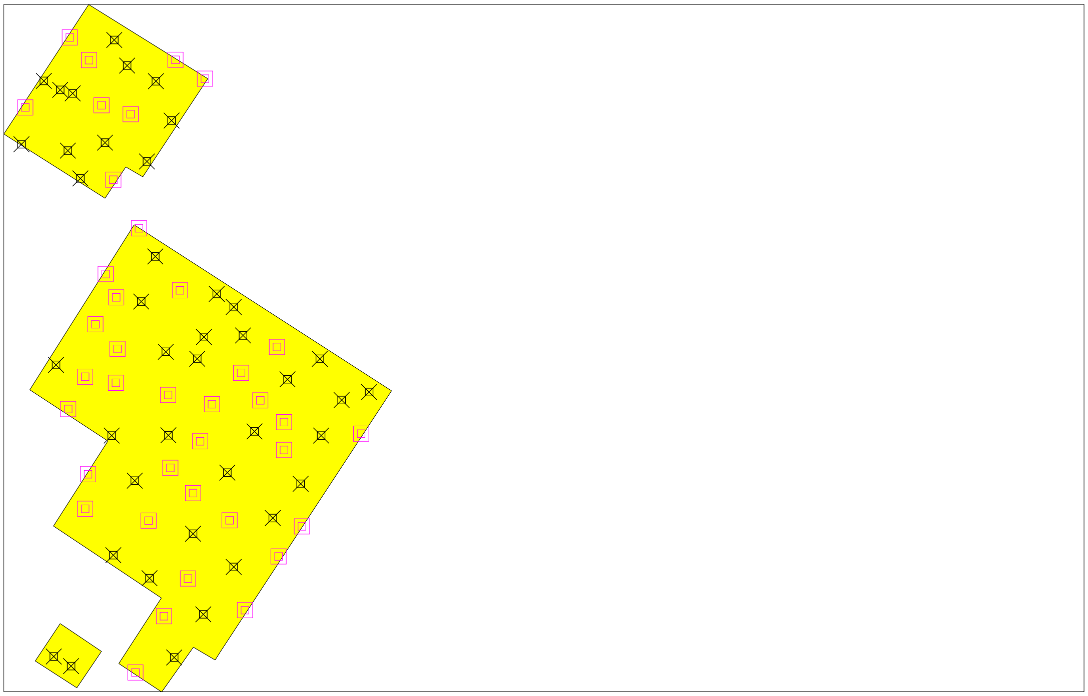
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	79
Number of samples on map [a]	79
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$40,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679242.7260	3083326.5280	J-07S	0.47	Manual	T
679104.2450	3083223.2620	J-02S	0.34	Manual	T
679181.2750	3083178.2880	J-08S	0.29	Manual	T
679133.4290	3083306.3130	J-01S	0.14	Manual	T
679268.7700	3083200.3260	J-11S	0.26	Manual	T
679171.2970	3083289.7960	J-04S	0.115	Manual	T
679164.8060	3083214.7100	J-06S	0.23	Manual	T
679213.7730	3083224.9730	J-09S	0.105	Manual	T
679280.5440	3083305.6810	J-10S	0.105	Manual	T
679155.0740	3083294.6960	J-03S	0.1	Manual	T
679301.1600	3083254.0340	J-12S	0.09	Manual	T
679225.8560	3083359.9740	J-05S	0.08	Manual	T
679344.9917	3083309.2278	J-38S	2.2	Adaptive-Fill	
679167.5023	3083363.3712	J-37S	2.1	Adaptive-Fill	
679247.5039	3083262.4916	J-19S	1.7	Adaptive-Fill	
679108.9961	3083271.3916	J-36S	1.7	Adaptive-Fill	
679224.6687	3083176.2417	J-41S	1.7	Adaptive-Fill	
679192.8634	3083333.4953	J-23S	1.4	Adaptive-Fill	
679208.9664	3083274.1678	J-28S	1.2	Adaptive-Fill	
679306.1204	3083334.0306	J-32S	1.1	Adaptive-Fill	

Area: Area 3

X Coord	Y Coord	Label	Value	Type	Historical
679470.3570	3082776.7350	J-38S	2.2	Manual	T
679433.9450	3082731.6820	J-37S	2.1	Manual	T
679343.5810	3082969.5980	J-19S	1.7	Manual	T
679382.8900	3082667.5270	J-36S	1.7	Manual	T
679560.6070	3082897.2580	J-41S	1.7	Manual	T
679297.0010	3082840.6970	J-23S	1.4	Manual	T
679272.0040	3082652.6750	J-28S	1.2	Manual	T
679453.4760	3082914.1150	J-32S	1.1	Manual	T
679146.6460	3082549.7640	J-25S	1	Manual	T
679169.0760	3082537.3510	J-27S	1	Manual	T
679495.8840	3082940.9730	J-33S	0.81	Manual	T
679304.6530	3082548.6880	J-34S	0.8	Manual	T
679342.7410	3082605.3190	J-35S	0.78	Manual	T
679252.7130	3082781.0290	J-22S	0.74	Manual	T
679497.3310	3082840.3960	J-39S	0.7	Manual	T
679524.3310	3082886.8990	J-40S	0.62	Manual	T
679329.4380	3082711.0960	J-29S	0.58	Manual	T
679149.4920	3082933.0980	J-13S	0.54	Manual	T
679374.4420	3082791.3300	J-30S	0.54	Manual	T
679360.5700	3083026.4980	J-18S	0.44	Manual	T
679293.5600	3082950.4980	J-17S	0.41	Manual	T
679261.0980	3083016.3510	J-15S	0.38	Manual	T
679222.6340	3082840.1720	J-16S	0.33	Manual	T
679394.8070	3082971.8300	J-24S	0.3	Manual	T
679382.8640	3083009.1130	J-20S	0.12	Manual	T
679279.6830	3083075.4290	J-14S	0.115	Manual	T
679410.1490	3082845.8460	J-31S	0.23	Manual	T
679224.5850	3082683.1400	J-26S	0.11	Manual	T
679335.0020	3082941.1720	J-21S	0.1	Manual	T
679187.7108	3082743.9049		0	Adaptive-Fill	
679201.3419	3082986.3558		0	Adaptive-Fill	
679228.1261	3082909.4864		0	Adaptive-Fill	
679354.1234	3082881.5703		0	Adaptive-Fill	
679165.3089	3082875.0841		0	Adaptive-Fill	
679214.7953	3083052.5423		0	Adaptive-Fill	
679270.7295	3082728.2056		0	Adaptive-Fill	
679253.6530	3082528.8006		0	Adaptive-Fill	
679296.7054	3082893.3847		0	Adaptive-Fill	
679329.1669	3082764.4865		0	Adaptive-Fill	
679322.5779	3082652.2311		0	Adaptive-Fill	

679377.2589	3082728.9013		0	Adaptive-Fill	
679397.3161	3082610.7954		0	Adaptive-Fill	
679392.4419	3082922.3762		0	Adaptive-Fill	
679291.0689	3082602.5607		0	Adaptive-Fill	
679191.3863	3082789.3978		0	Adaptive-Fill	
679439.5764	3082956.5943		0	Adaptive-Fill	
679441.5982	3082681.4673		0	Adaptive-Fill	
679448.5842	3082821.4334		0	Adaptive-Fill	
679299.4420	3082797.6478		0	Adaptive-Fill	
679549.8870	3082842.9025		0	Adaptive-Fill	
679312.1684	3083030.8840		0	Adaptive-Fill	
679230.0779	3082953.9913		0	Adaptive-Fill	
679338.6362	3082832.5643		0	Adaptive-Fill	
679472.1177	3082720.7150		0	Adaptive-Fill	
679258.2576	3083112.5898		0	Adaptive-Fill	
679417.6739	3082886.3804		0	Adaptive-Fill	
679187.6340	3082917.6887		0	Adaptive-Fill	
679448.6706	3082857.7165		0	Adaptive-Fill	
679228.4704	3083021.7000		0	Adaptive-Fill	

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

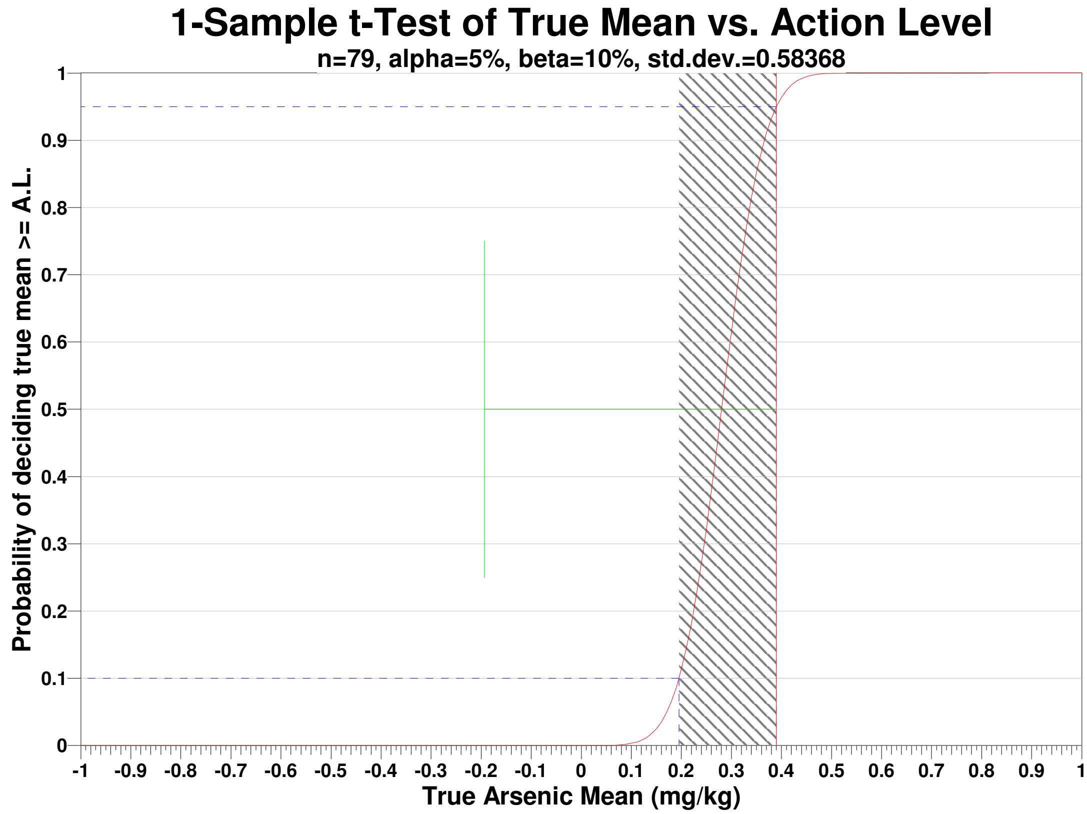
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Arsenic	79	0.58368 mg/kg	0.1948 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=0.39		α=5		α=10		α=15	
		s=1.16736	s=0.58368	s=1.16736	s=0.58368	s=1.16736	s=0.58368
LBGR=90	β=5	9698	2426	7674	1920	6442	1611
	β=10	7675	1920	5887	1473	4815	1205
	β=15	6443	1612	4815	1205	3851	963
LBGR=80	β=5	2426	608	1920	481	1611	404
	β=10	1920	481	1473	369	1205	302
	β=15	1612	404	1205	302	963	242
LBGR=70	β=5	1079	271	854	214	717	180

	β=10	854	215	655	165	536	135
	β=15	718	181	536	135	429	108

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $40,500.00, which averages out to a per sample cost of $512.66. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**79 Samples**
Field collection costs		$100.00	$7,900.00
Analytical costs	$400.00	$400.00	$31,600.00
Sum of Field & Analytical costs		**$500.00**	**$39,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$40,500.00**

Data Analysis for Arsenic
The following data points were entered by the user for analysis.

Arsenic (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0
30	0.08	0.09	0.1	0.1	0.105	0.105	0.11	0.115	0.115	0.12
40	0.14	0.23	0.23	0.26	0.29	0.3	0.33	0.34	0.38	0.41
50	0.44	0.47	0.54	0.54	0.58	0.62	0.7	0.74	0.78	0.8
60	0.81	1	1	1.1	1.1	1.2	1.2	1.4	1.4	1.7
70	1.7	1.7	1.7	1.7	1.7	2.1	2.1	2.2	2.2	

SUMMARY STATISTICS for Arsenic	
n	79
Min	0
Max	2.2
Range	2.2
Mean	0.49582
Median	0.12
Variance	0.42709

StdDev				0.65352				
Std Error				0.073527				
Skewness				1.2833				
Interquartile Range				0.8				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0	0	0	0	0.12	0.8	1.7	2.1	2.2

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Arsenic			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	2.68	3.05	No

None of the test statistics exceeded the corresponding critical values, therefore none of the 1 tests are significant and we conclude that at the 5% significance level there are no outliers in the data.

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Lilliefors test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Lilliefors Test Statistic	0.8461
Lilliefors 5% Critical Value	0.94

The calculated Lilliefors test statistic is less than the 5% Lilliefors critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

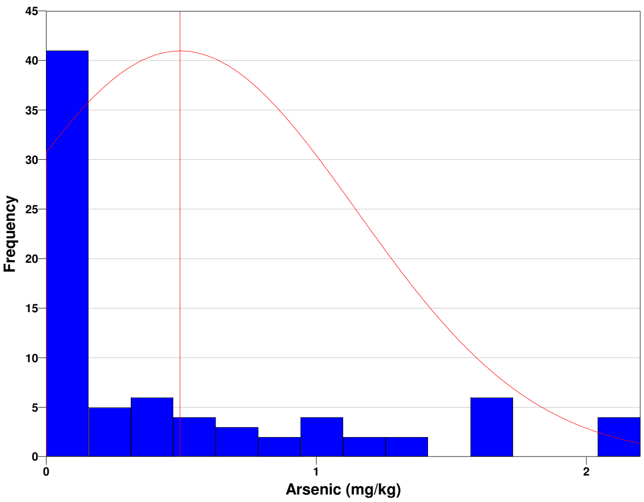
Data Plots for Arsenic
Graphical displays of the data are shown below.

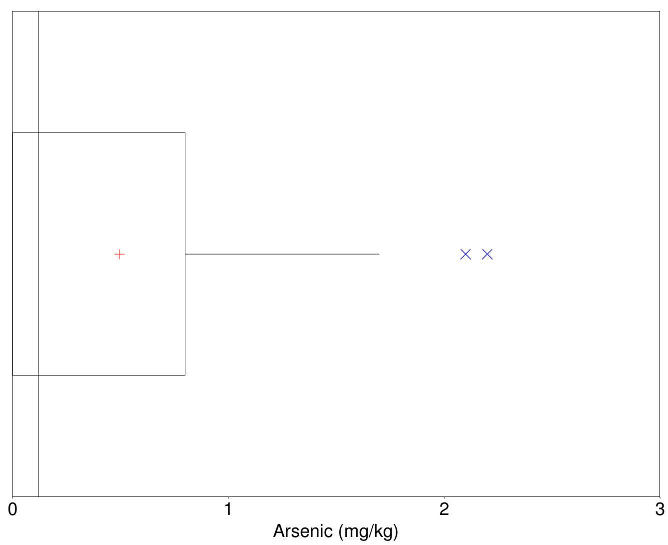
The Histogram is a plot of the fraction of the n observed data that fall within specified data "bins." A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

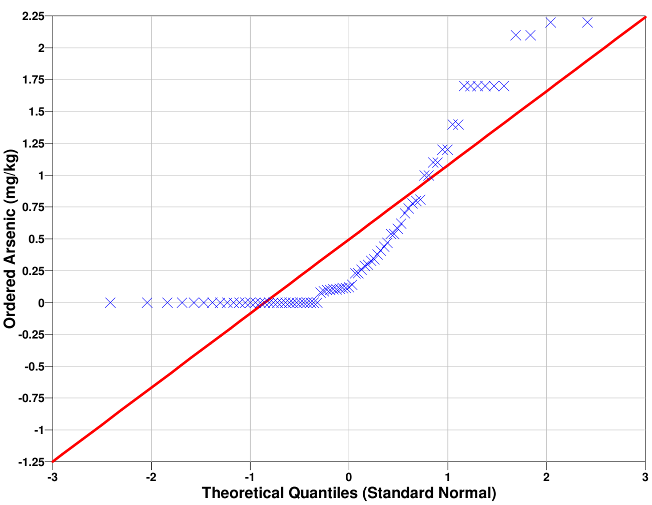
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Arsenic
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Lilliefors test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Lilliefors Test Statistic	0.2259
Lilliefors 5% Critical Value	0.09968

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.6182
95% Non-Parametric (Chebyshev) UCL	0.8163

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.8163) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=79 data,
AL is the action level or threshold (0.39),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=78 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
1.4392	1.6646	Cannot Reject

The test did not reject the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean exceeds the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
49	47	Reject

Random sampling locations for comparing a mean with a fixed threshold (parametric)

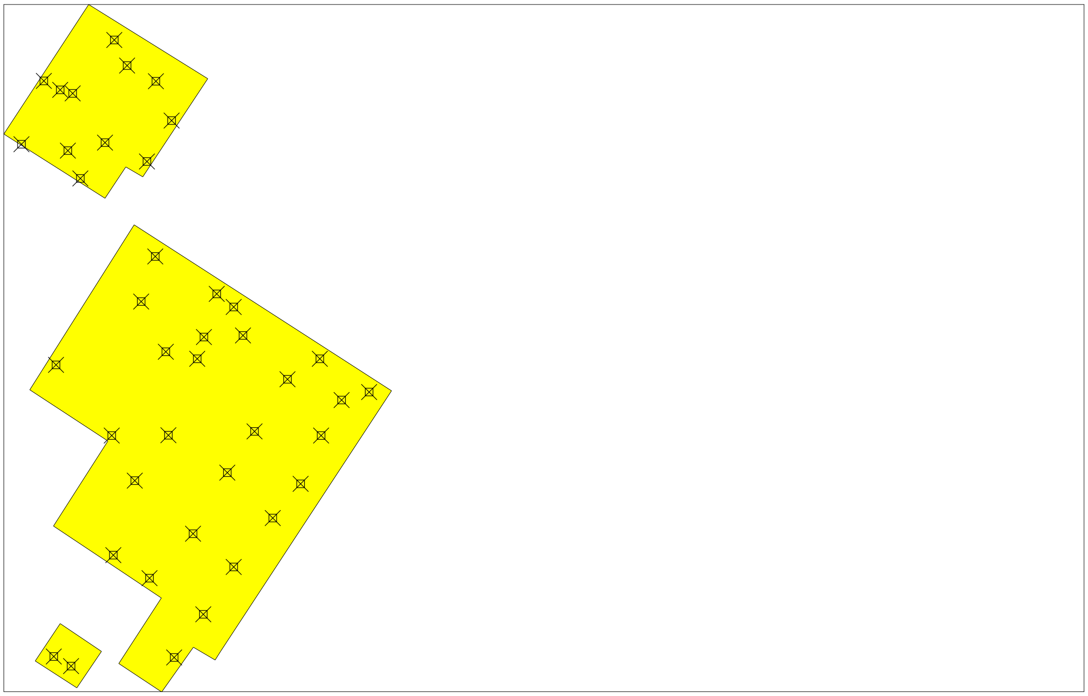
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679242.7260	3083326.5280	J-07S	0.47	Manual	T
679104.2450	3083223.2620	J-02S	0.34	Manual	T
679181.2750	3083178.2880	J-08S	0.29	Manual	T
679133.4290	3083306.3130	J-01S	0.14	Manual	T
679268.7700	3083200.3260	J-11S	0.26	Manual	T
679171.2970	3083289.7960	J-04S	0.115	Manual	T
679164.8060	3083214.7100	J-06S	0.23	Manual	T
679213.7730	3083224.9730	J-09S	0.105	Manual	T
679280.5440	3083305.6810	J-10S	0.105	Manual	T
679155.0740	3083294.6960	J-03S	0.1	Manual	T
679301.1600	3083254.0340	J-12S	0.09	Manual	T
679225.8560	3083359.9740	J-05S	0.08	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679470.3570	3082776.7350	J-38S	2.2	Manual	T
679433.9450	3082731.6820	J-37S	2.1	Manual	T
679343.5810	3082969.5980	J-19S	1.7	Manual	T
679382.8900	3082667.5270	J-36S	1.7	Manual	T
679560.6070	3082897.2580	J-41S	1.7	Manual	T
679297.0010	3082840.6970	J-23S	1.4	Manual	T
679272.0040	3082652.6750	J-28S	1.2	Manual	T

679453.4760	3082914.1150	J-32S	1.1	Manual	T
679146.6460	3082549.7640	J-25S	1	Manual	T
679169.0760	3082537.3510	J-27S	1	Manual	T
679495.8840	3082940.9730	J-33S	0.81	Manual	T
679304.6530	3082548.6880	J-34S	0.8	Manual	T
679342.7410	3082605.3190	J-35S	0.78	Manual	T
679252.7130	3082781.0290	J-22S	0.74	Manual	T
679497.3310	3082840.3960	J-39S	0.7	Manual	T
679524.3310	3082886.8990	J-40S	0.62	Manual	T
679329.4380	3082711.0960	J-29S	0.58	Manual	T
679149.4920	3082933.0980	J-13S	0.54	Manual	T
679374.4420	3082791.3300	J-30S	0.54	Manual	T
679360.5700	3083026.4980	J-18S	0.44	Manual	T
679293.5600	3082950.4980	J-17S	0.41	Manual	T
679261.0980	3083016.3510	J-15S	0.38	Manual	T
679222.6340	3082840.1720	J-16S	0.33	Manual	T
679394.8070	3082971.8300	J-24S	0.3	Manual	T
679382.8640	3083009.1130	J-20S	0.12	Manual	T
679279.6830	3083075.4290	J-14S	0.115	Manual	T
679410.1490	3082845.8460	J-31S	0.23	Manual	T
679224.5850	3082683.1400	J-26S	0.11	Manual	T
679335.0020	3082941.1720	J-21S	0.1	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

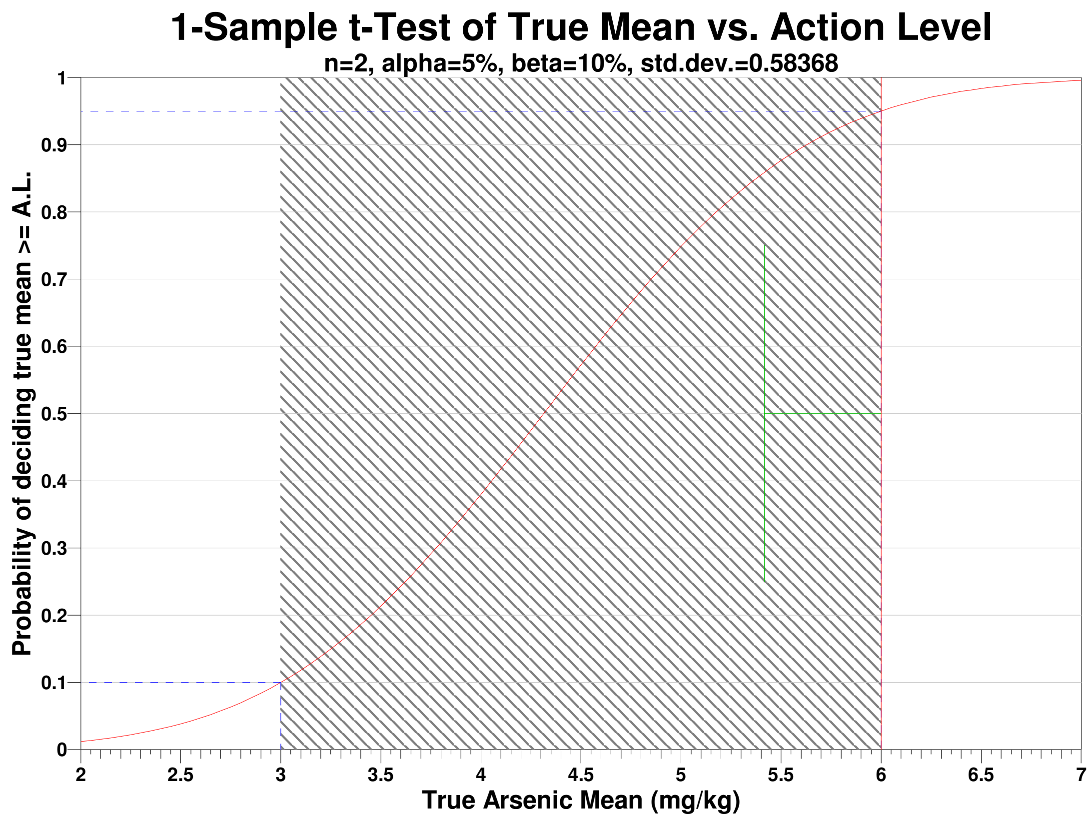
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Arsenic	2	0.58368 mg/kg	3 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=6		α=5		α=10		α=15	
		s=1.16736	s=0.58368	s=1.16736	s=0.58368	s=1.16736	s=0.58368
LBGR=90	β=5	43	12	34	9	28	8
	β=10	34	10	26	8	21	6
	β=15	29	9	22	6	17	5
LBGR=80	β=5	12	4	9	3	8	3
	β=10	10	4	8	3	6	2
	β=15	9	4	6	3	5	2
LBGR=70	β=5	6	3	5	2	4	2

	β=10	5	3	4	2	3	2
	β=15	5	3	4	2	3	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Arsenic
The following data points were entered by the user for analysis.

Arsenic (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0
30	0.08	0.09	0.1	0.1	0.105	0.105	0.11	0.115	0.115	0.12
40	0.14	0.23	0.23	0.26	0.29	0.3	0.33	0.34	0.38	0.41
50	0.44	0.47	0.54	0.54	0.58	0.62	0.7	0.74	0.78	0.8
60	0.81	1	1	1.1	1.1	1.2	1.2	1.4	1.4	1.7
70	1.7	1.7	1.7	1.7	1.7	2.1	2.1	2.2	2.2	

SUMMARY STATISTICS for Arsenic	
n	79
Min	0
Max	2.2
Range	2.2
Mean	0.49582
Median	0.12
Variance	0.42709

StdDev				0.65352				
Std Error				0.073527				
Skewness				1.2833				
Interquartile Range				0.8				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0	0	0	0	0.12	0.8	1.7	2.1	2.2

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Arsenic			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	2.608	3.305	No

None of the test statistics exceeded the corresponding critical values, therefore none of the 1 tests are significant and we conclude that at the 5% significance level there are no outliers in the data.

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Lilliefors test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Lilliefors Test Statistic	0.2282
Lilliefors 5% Critical Value	0.1003

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

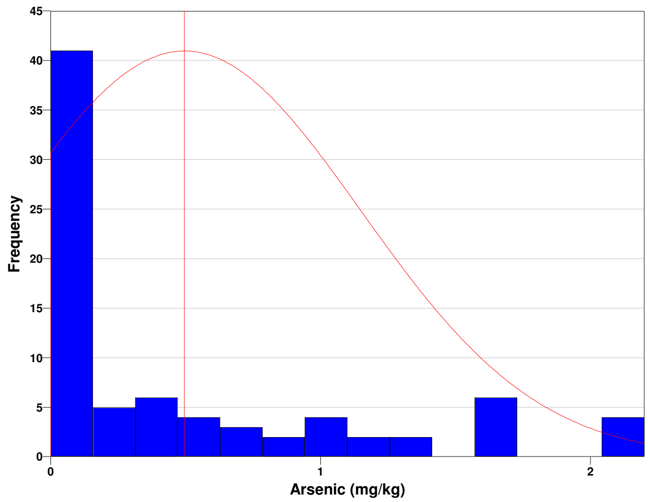
Data Plots for Arsenic
Graphical displays of the data are shown below.

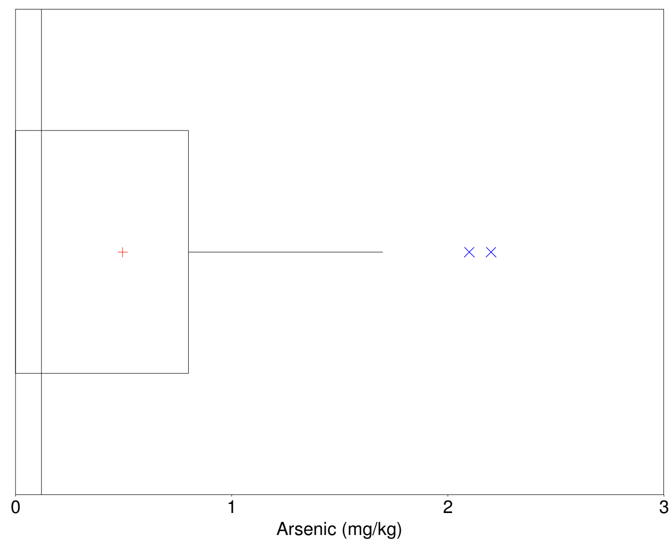
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

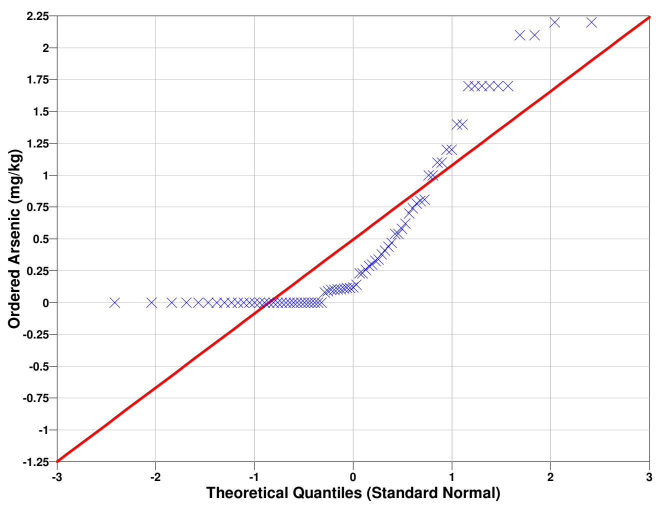
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Arsenic
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Lilliefors test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Lilliefors Test Statistic	0.2259
Lilliefors 5% Critical Value	0.09968

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean

Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.6182
95% Non-Parametric (Chebyshev) UCL	0.8163

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.8163) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test

A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where

x is the sample mean of the n=79 data,

AL is the action level or threshold (6),

SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=78 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-74.86	1.6646	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
79	47	Reject

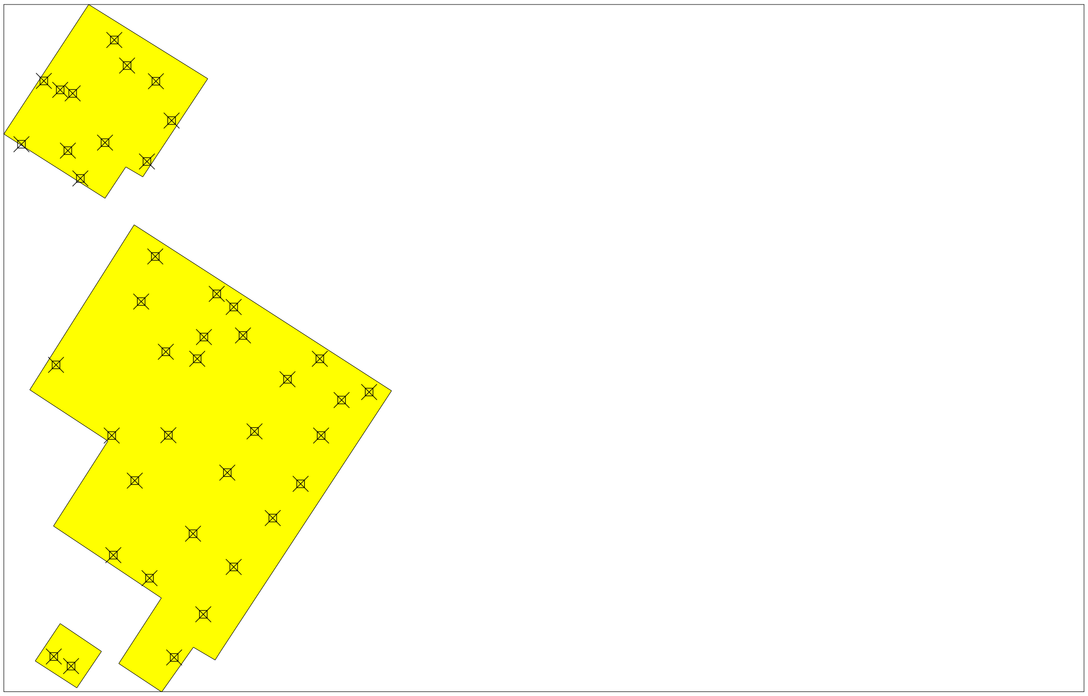
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.98	Manual	T
679301.1600	3083254.0340	J-12S	0.58	Manual	T
679171.2970	3083289.7960	J-04S	2.1	Manual	T
679155.0740	3083294.6960	J-03S	1.2	Manual	T
679133.4290	3083306.3130	J-01S	0.63	Manual	T
679104.2450	3083223.2620	J-02S	3.9	Manual	T
679164.8060	3083214.7100	J-06S	3.7	Manual	T
679181.2750	3083178.2880	J-08S	5.2	Manual	T
679213.7730	3083224.9730	J-09S	1.8	Manual	T
679280.5440	3083305.6810	J-10S	1	Manual	T
679242.7260	3083326.5280	J-07S	4.3	Manual	T
679225.8560	3083359.9740	J-05S	0.8	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	3.4	Manual	T
679335.0020	3082941.1720	J-21S	0.59	Manual	T
679343.5810	3082969.5980	J-19S	15	Manual	T
679394.8070	3082971.8300	J-24S	1.2	Manual	T
679382.8640	3083009.1130	J-20S	2.5	Manual	T
679293.5600	3082950.4980	J-17S	2.4	Manual	T
679360.5700	3083026.4980	J-18S	3.2	Manual	T

679297.0010	3082840.6970	J-23S	3.5	Manual	T
679252.7130	3082781.0290	J-22S	4.1	Manual	T
679222.6340	3082840.1720	J-16S	3.2	Manual	T
679279.6830	3083075.4290	J-14S	0.6	Manual	T
679149.4920	3082933.0980	J-13S	4.9	Manual	T
679272.0040	3082652.6750	J-28S	6.1	Manual	T
679224.5850	3082683.1400	J-26S	1.7	Manual	T
679146.6460	3082549.7640	J-25S	7.4	Manual	T
679169.0760	3082537.3510	J-27S	3.7	Manual	T
679342.7410	3082605.3190	J-35S	1.8	Manual	T
679304.6530	3082548.6880	J-34S	1.5	Manual	T
679382.8900	3082667.5270	J-36S	6	Manual	T
679329.4380	3082711.0960	J-29S	2.7	Manual	T
679374.4420	3082791.3300	J-30S	2.2	Manual	T
679453.4760	3082914.1150	J-32S	2.9	Manual	T
679410.1490	3082845.8460	J-31S	0.76	Manual	T
679560.6070	3082897.2580	J-41S	2.2	Manual	T
679495.8840	3082940.9730	J-33S	1.9	Manual	T
679524.3310	3082886.8990	J-40S	2.3	Manual	T
679497.3310	3082840.3960	J-39S	1.4	Manual	T
679470.3570	3082776.7350	J-38S	2.5	Manual	T
679433.9450	3082731.6820	J-37S	8.8	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

- n is the number of samples,
- S is the estimated standard deviation of the measured values including analytical error,
- Δ is the width of the gray region,
- α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
- β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
- $Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
- $Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Chromium	2	2.69 mg/kg	105.338 mg/kg	0.05	0.1	1.64485	1.28155

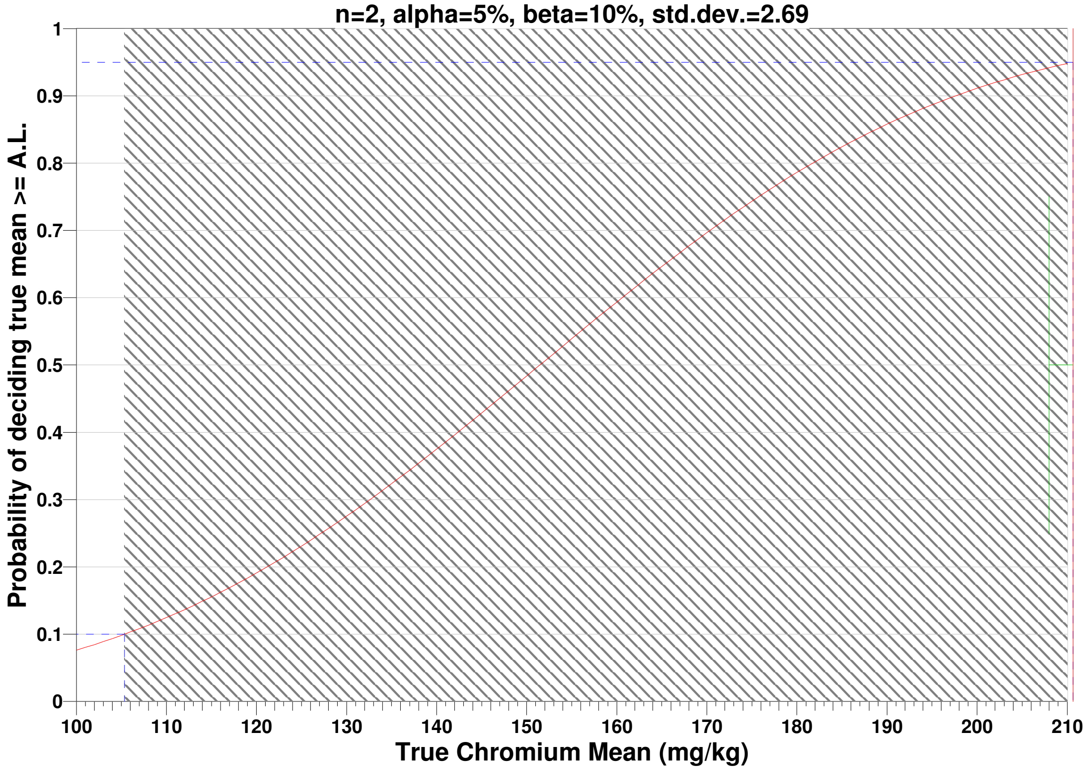
[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.

1-Sample t-Test of True Mean vs. Action Level

n=2, alpha=5%, beta=10%, std.dev.=2.69



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=210.675		α=5		α=10		α=15	
		s=5.38	s=2.69	s=5.38	s=2.69	s=5.38	s=2.69
LBGR=90	β=5	3	2	2	1	2	1
	β=10	2	2	2	1	1	1
	β=15	2	2	2	1	1	1
LBGR=80	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	β=5	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Chromium
The following data points were entered by the user for analysis.

Chromium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.58	0.59	0.6	0.63	0.76	0.8	0.98	1	1.2	1.2
10	1.4	1.5	1.7	1.8	1.8	1.9	2.1	2.2	2.2	2.3
20	2.4	2.5	2.5	2.7	2.9	3.2	3.2	3.4	3.5	3.7
30	3.7	3.9	4.1	4.3	4.9	5.2	6	6.1	7.4	8.8
40	15									

SUMMARY STATISTICS for Chromium	
n	41
Min	0.58
Max	15
Range	14.42
Mean	3.0888
Median	2.4
Variance	7.2627
StdDev	2.6949
Std Error	0.42088
Skewness	2.5355

Interquartile Range				2.5				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.58	0.591	0.656	1.3	2.4	3.8	6.08	8.66	15

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Chromium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	4.42	3.05	Yes

The test statistic 4.42 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Chromium	
1	15

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.8953
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

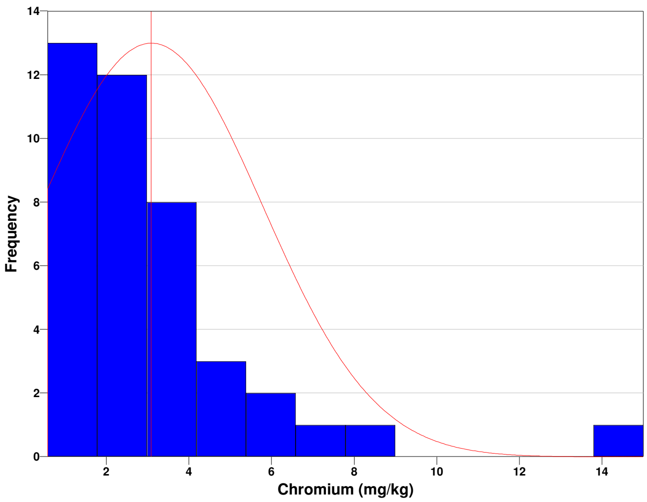
Data Plots for Chromium
Graphical displays of the data are shown below.

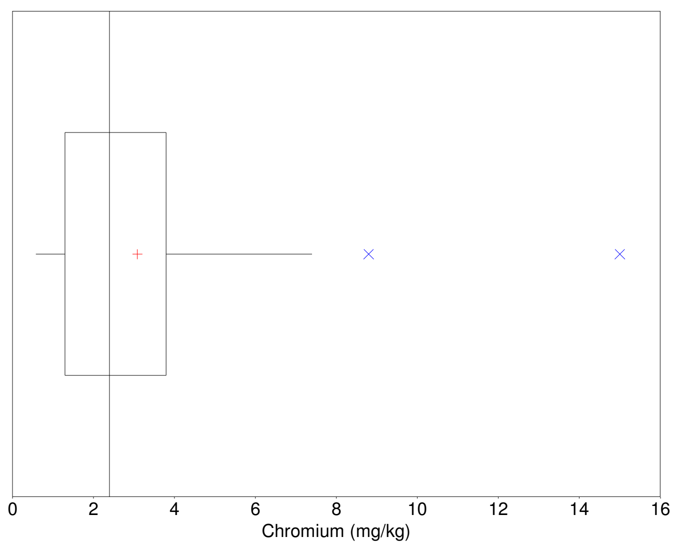
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

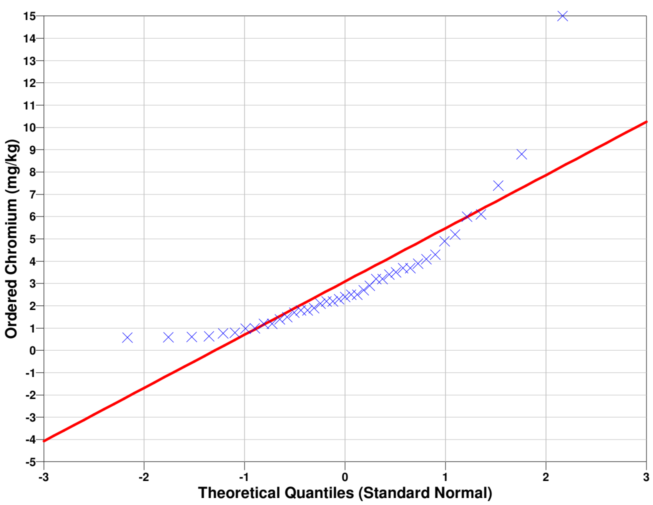
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Chromium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.7725
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	3.797
95% Non-Parametric (Chebyshev) UCL	4.923

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (4.923) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (210.675),
SE is the standard error = (standard deviation) / (square root of n).

This *t* was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-493.22	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

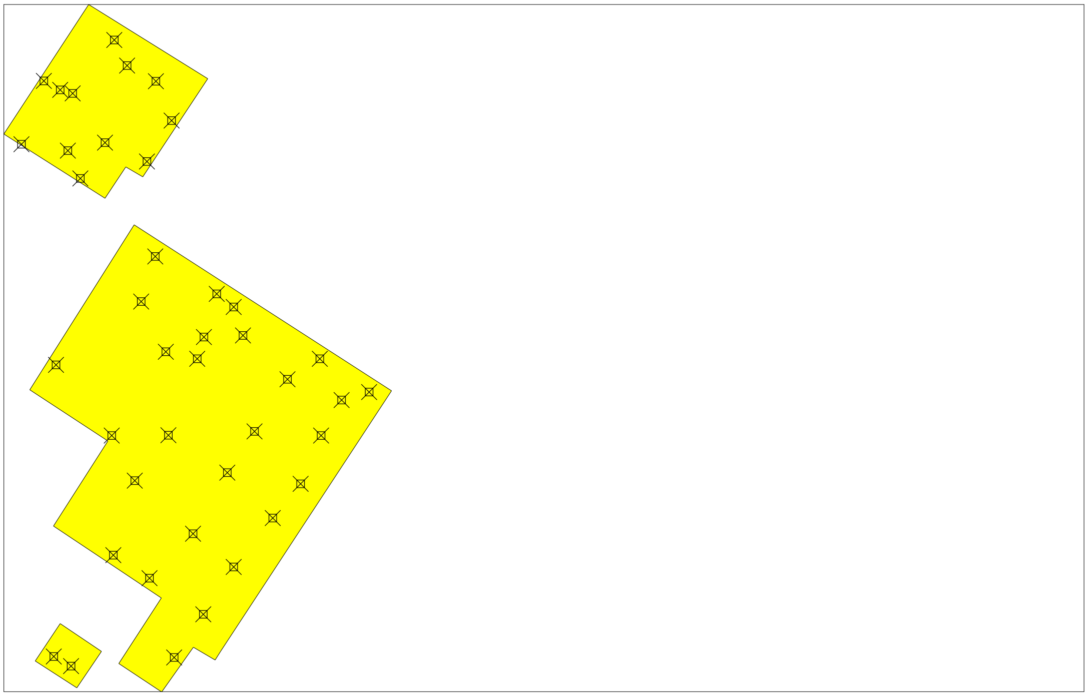
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.98	Manual	T
679301.1600	3083254.0340	J-12S	0.58	Manual	T
679171.2970	3083289.7960	J-04S	2.1	Manual	T
679155.0740	3083294.6960	J-03S	1.2	Manual	T
679133.4290	3083306.3130	J-01S	0.63	Manual	T
679104.2450	3083223.2620	J-02S	3.9	Manual	T
679164.8060	3083214.7100	J-06S	3.7	Manual	T
679181.2750	3083178.2880	J-08S	5.2	Manual	T
679213.7730	3083224.9730	J-09S	1.8	Manual	T
679280.5440	3083305.6810	J-10S	1	Manual	T
679242.7260	3083326.5280	J-07S	4.3	Manual	T
679225.8560	3083359.9740	J-05S	0.8	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	3.4	Manual	T
679335.0020	3082941.1720	J-21S	0.59	Manual	T
679343.5810	3082969.5980	J-19S	15	Manual	T
679394.8070	3082971.8300	J-24S	1.2	Manual	T
679382.8640	3083009.1130	J-20S	2.5	Manual	T
679293.5600	3082950.4980	J-17S	2.4	Manual	T
679360.5700	3083026.4980	J-18S	3.2	Manual	T

679297.0010	3082840.6970	J-23S	3.5	Manual	T
679252.7130	3082781.0290	J-22S	4.1	Manual	T
679222.6340	3082840.1720	J-16S	3.2	Manual	T
679279.6830	3083075.4290	J-14S	0.6	Manual	T
679149.4920	3082933.0980	J-13S	4.9	Manual	T
679272.0040	3082652.6750	J-28S	6.1	Manual	T
679224.5850	3082683.1400	J-26S	1.7	Manual	T
679146.6460	3082549.7640	J-25S	7.4	Manual	T
679169.0760	3082537.3510	J-27S	3.7	Manual	T
679342.7410	3082605.3190	J-35S	1.8	Manual	T
679304.6530	3082548.6880	J-34S	1.5	Manual	T
679382.8900	3082667.5270	J-36S	6	Manual	T
679329.4380	3082711.0960	J-29S	2.7	Manual	T
679374.4420	3082791.3300	J-30S	2.2	Manual	T
679453.4760	3082914.1150	J-32S	2.9	Manual	T
679410.1490	3082845.8460	J-31S	0.76	Manual	T
679560.6070	3082897.2580	J-41S	2.2	Manual	T
679495.8840	3082940.9730	J-33S	1.9	Manual	T
679524.3310	3082886.8990	J-40S	2.3	Manual	T
679497.3310	3082840.3960	J-39S	1.4	Manual	T
679470.3570	3082776.7350	J-38S	2.5	Manual	T
679433.9450	3082731.6820	J-37S	8.8	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

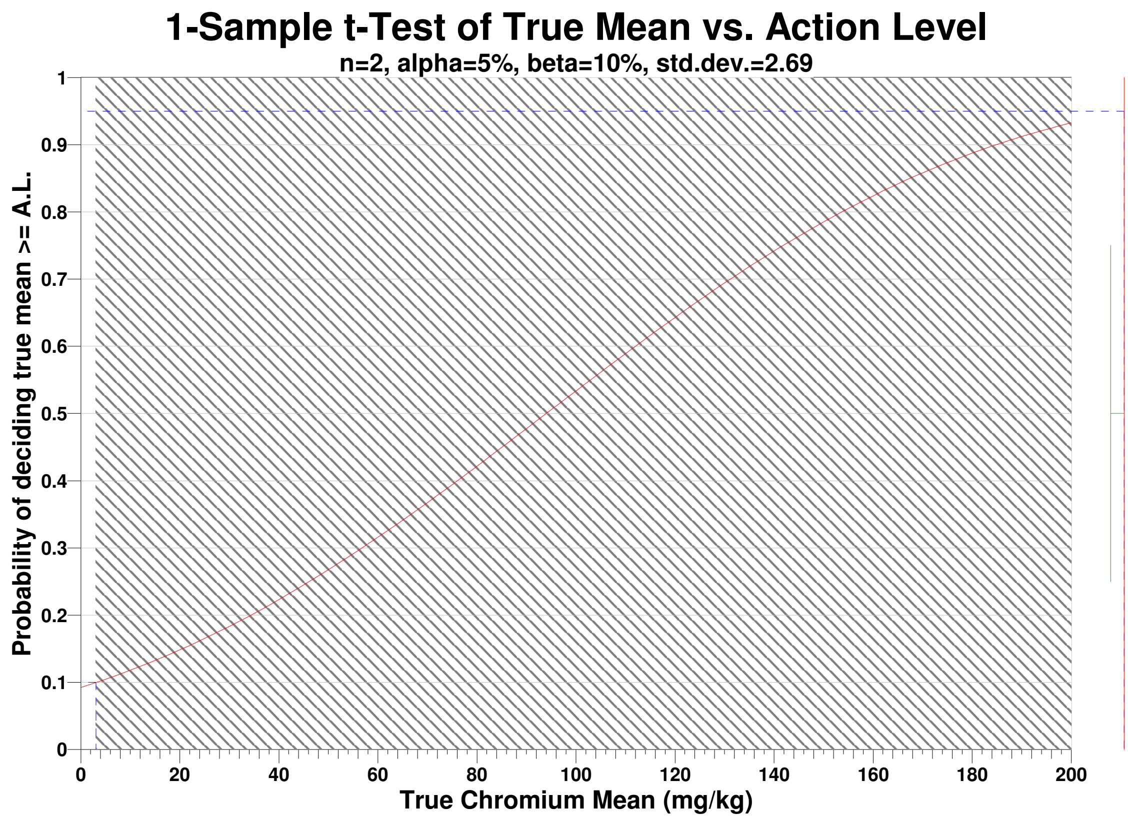
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	***$Z_{1-\alpha}$*** [a]	***$Z_{1-\beta}$*** [b]
Chromium	2	2.69 mg/kg	207.59 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=210.675		α=5		α=10		α=15	
		s=5.38	s=2.69	s=5.38	s=2.69	s=5.38	s=2.69
LBGR=90	β=5	3	2	2	1	2	1
	β=10	2	2	2	1	1	1
	β=15	2	2	2	1	1	1
LBGR=80	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	β=5	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Chromium
The following data points were entered by the user for analysis.

Chromium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.58	0.59	0.6	0.63	0.76	0.8	0.98	1	1.2	1.2
10	1.4	1.5	1.7	1.8	1.8	1.9	2.1	2.2	2.2	2.3
20	2.4	2.5	2.5	2.7	2.9	3.2	3.2	3.4	3.5	3.7
30	3.7	3.9	4.1	4.3	4.9	5.2	6	6.1	7.4	8.8
40	15									

SUMMARY STATISTICS for Chromium	
n	41
Min	0.58
Max	15
Range	14.42
Mean	3.0888
Median	2.4
Variance	7.2627
StdDev	2.6949
Std Error	0.42088
Skewness	2.5355

Interquartile Range				2.5				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.58	0.591	0.656	1.3	2.4	3.8	6.08	8.66	15

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Chromium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	4.42	3.05	Yes

The test statistic 4.42 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Chromium	
1	15

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.8953
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

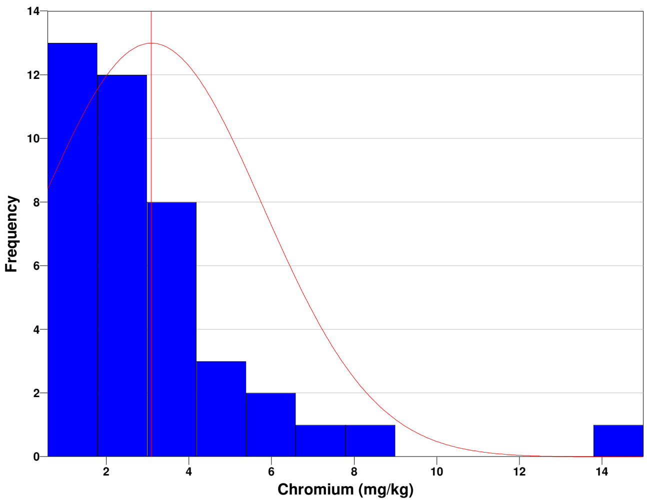
Data Plots for Chromium
Graphical displays of the data are shown below.

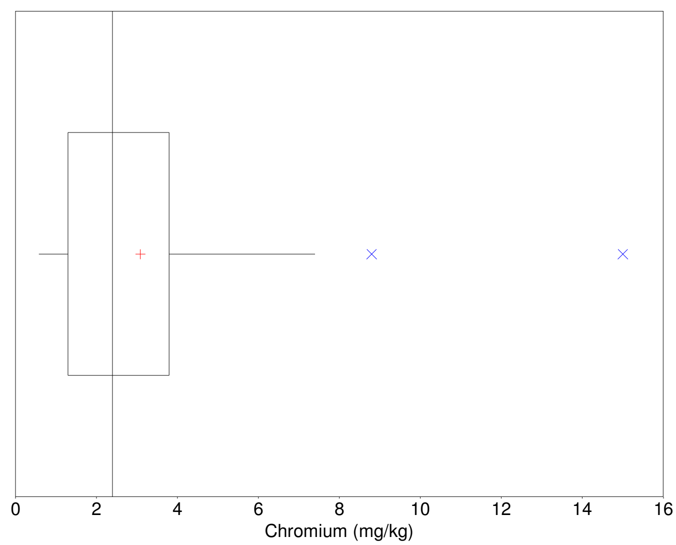
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

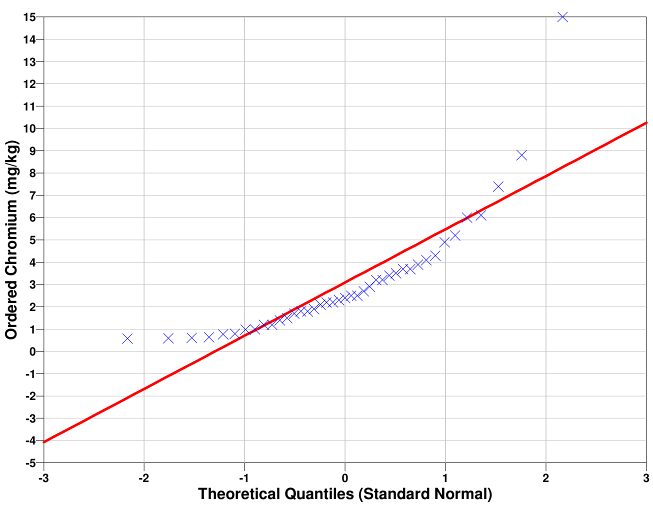
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The pth quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Chromium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.7725
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean

Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	3.797
95% Non-Parametric (Chebyshev) UCL	4.923

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (4.923) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test

A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where

x is the sample mean of the n=41 data,
AL is the action level or threshold (210.675),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-493.22	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

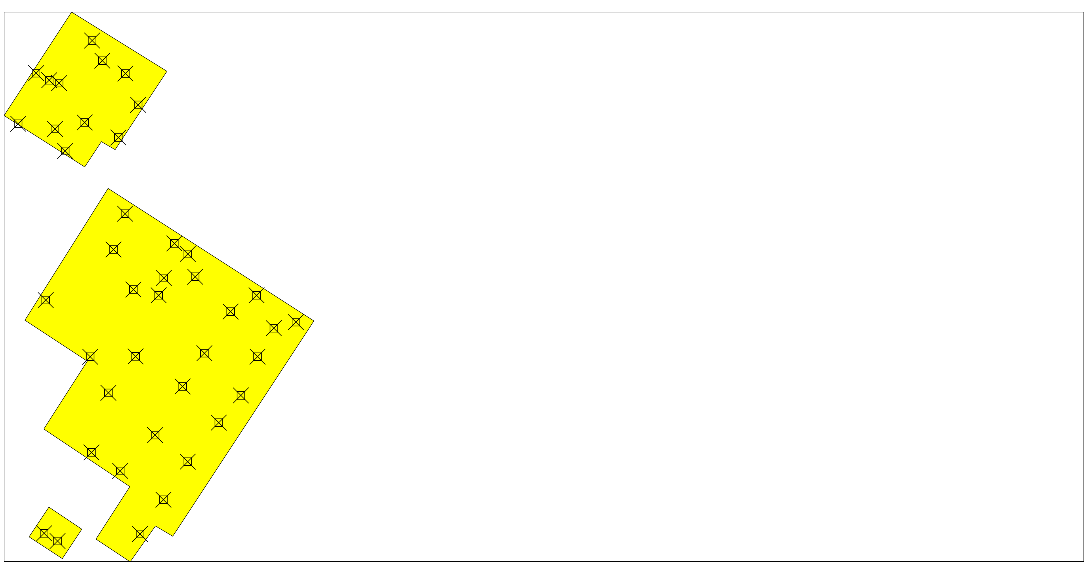
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1

X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.0017	Manual	T
679301.1600	3083254.0340	J-12S	0.59	Manual	T
679171.2970	3083289.7960	J-04S	0.00043	Manual	T
679155.0740	3083294.6960	J-03S	0.0051	Manual	T
679133.4290	3083306.3130	J-01S	0.055	Manual	T
679104.2450	3083223.2620	J-02S	0.019	Manual	T
679164.8060	3083214.7100	J-06S	0.033	Manual	T
679181.2750	3083178.2880	J-08S	0.048	Manual	T
679213.7730	3083224.9730	J-09S	0.00036	Manual	T
679280.5440	3083305.6810	J-10S	0.00038	Manual	T
679242.7260	3083326.5280	J-07S	0.012	Manual	T
679225.8560	3083359.9740	J-05S	0.0021	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	0.013	Manual	T
679335.0020	3082941.1720	J-21S	0.00038	Manual	T
679343.5810	3082969.5980	J-19S	0.0065	Manual	T
679394.8070	3082971.8300	J-24S	0.012	Manual	T
679382.8640	3083009.1130	J-20S	0.00044	Manual	T
679293.5600	3082950.4980	J-17S	0.055	Manual	T
679360.5700	3083026.4980	J-18S	0.0048	Manual	T
679297.0010	3082840.6970	J-23S	0.0077	Manual	T
679252.7130	3082781.0290	J-22S	0.0045	Manual	T
679222.6340	3082840.1720	J-16S	0.0025	Manual	T
679279.6830	3083075.4290	J-14S	0.00038	Manual	T
679149.4920	3082933.0980	J-13S	0.0073	Manual	T
679272.0040	3082652.6750	J-28S	0.054	Manual	T
679224.5850	3082683.1400	J-26S	0.011	Manual	T
679146.6460	3082549.7640	J-25S	0.01	Manual	T
679169.0760	3082537.3510	J-27S	0.0072	Manual	T
679342.7410	3082605.3190	J-35S	0.0038	Manual	T
679304.6530	3082548.6880	J-34S	0.000385	Manual	T
679382.8900	3082667.5270	J-36S	0.008	Manual	T
679329.4380	3082711.0960	J-29S	0.0024	Manual	T
679374.4420	3082791.3300	J-30S	0.0043	Manual	T
679453.4760	3082914.1150	J-32S	0.000365	Manual	T
679410.1490	3082845.8460	J-31S	0.0026	Manual	T
679560.6070	3082897.2580	J-41S	0.0053	Manual	T
679495.8840	3082940.9730	J-33S	0.0048	Manual	T

679524.3310	3082886.8990	J-40S	0.0046	Manual	T
679497.3310	3082840.3960	J-39S	0.014	Manual	T
679470.3570	3082776.7350	J-38S	0.0013	Manual	T
679433.9450	3082731.6820	J-37S	0.00035	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

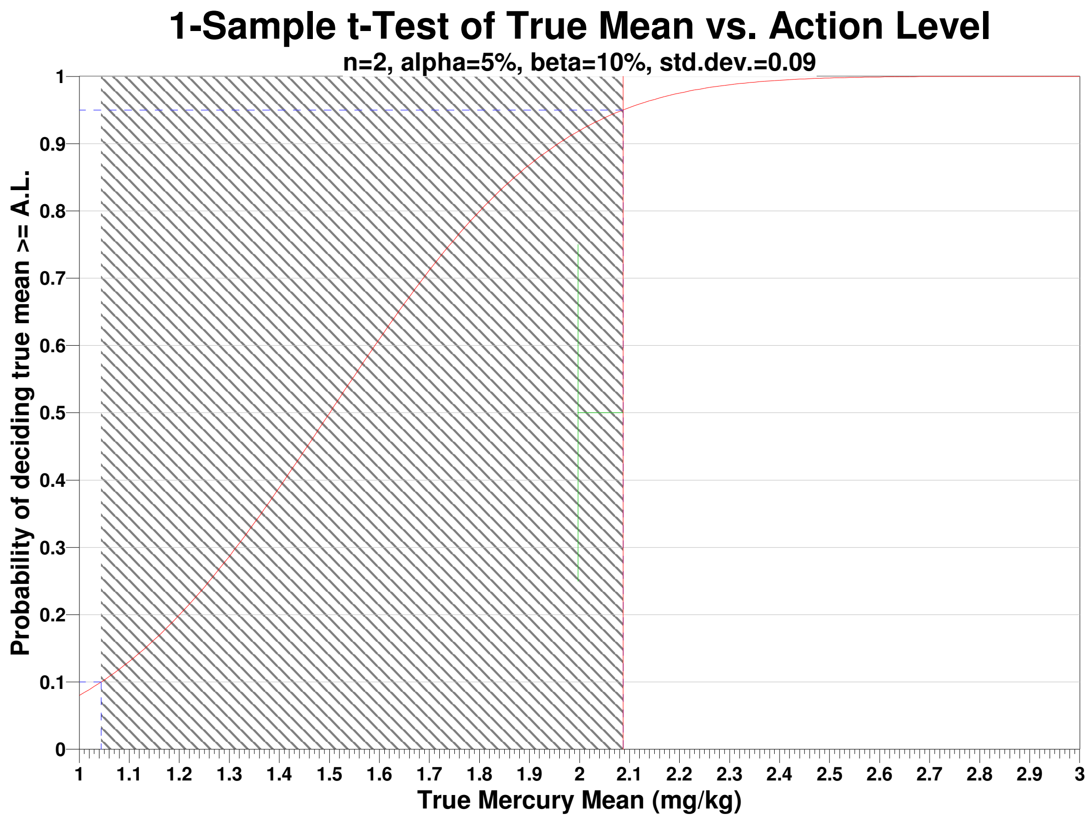
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	Δ	α	β	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Mercury	2	0.09 mg/kg	1.0436 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples			
AL=2.0872	α=5	α=10	α=15

		s=0.18	s=0.09	s=0.18	s=0.09	s=0.18	s=0.09
LBGR=90	**β=5**	10	4	8	3	6	2
	β=10	8	3	6	3	5	2
	β=15	7	3	5	2	4	2
LBGR=80	**β=5**	4	2	3	2	2	1
	β=10	3	2	3	2	2	1
	β=15	3	2	2	2	2	1
LBGR=70	**β=5**	3	2	2	1	2	1
	β=10	3	2	2	1	1	1
	β=15	2	2	2	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Mercury
The following data points were entered by the user for analysis.

Mercury (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.00035	0.00036	0.000365	0.00038	0.00038	0.00038	0.000385	0.00043	0.00044	0.0013
10	0.0017	0.0021	0.0024	0.0025	0.0026	0.0038	0.0043	0.0045	0.0046	0.0048
20	0.0048	0.0051	0.0053	0.0065	0.0072	0.0073	0.0077	0.008	0.01	0.011
30	0.012	0.012	0.013	0.014	0.019	0.033	0.048	0.054	0.055	0.055
40	0.59									

SUMMARY STATISTICS for Mercury	
n	41
Min	0.00035

Max				0.59				
Range				0.58965				
Mean				0.02478				
Median				0.0048				
Variance				0.0084248				
StdDev				0.091787				
Std Error				0.014335				
Skewness				6.13				
Interquartile Range				0.0105				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.00035	0.0003605	0.00038	0.0015	0.0048	0.012	0.0528	0.055	0.59

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Mercury			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	6.158	3.05	Yes

The test statistic 6.158 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Mercury	
1	0.59

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.6364
Shapiro-Wilk 5% Critical Value	0.94

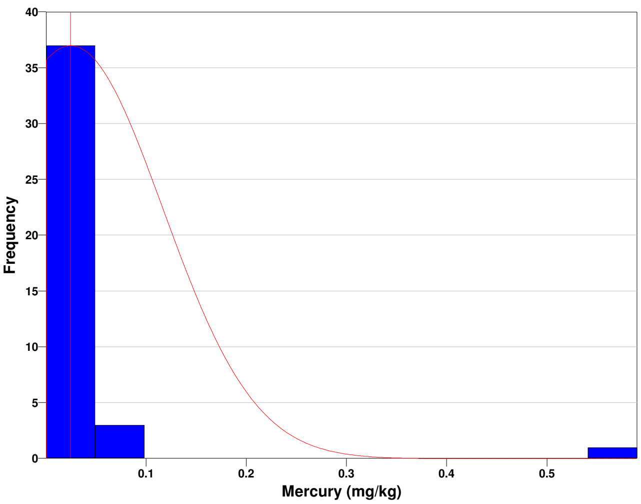
The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

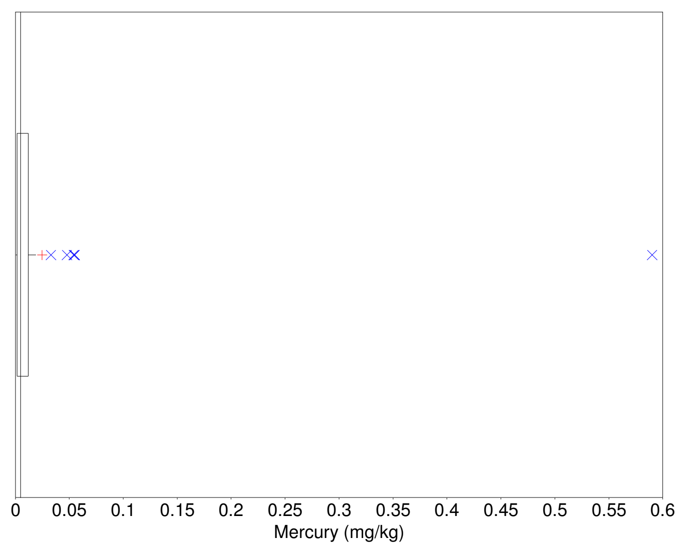
Data Plots for Mercury
Graphical displays of the data are shown below.

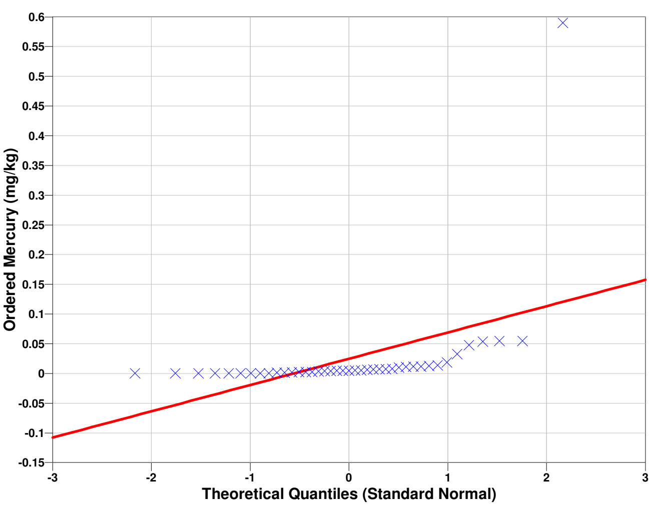
The Histogram is a plot of the fraction of the n observed data that fall within specified data "bins." A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25th and 75th percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The pth quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through

2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Mercury
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.2612
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.04892
95% Non-Parametric (Chebyshev) UCL	0.08726

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.08726) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (2.0872),
SE is the standard error = (standard deviation) / (square root of n).

This *t* was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-143.88	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

Random sampling locations for comparing a mean with a fixed threshold (parametric)

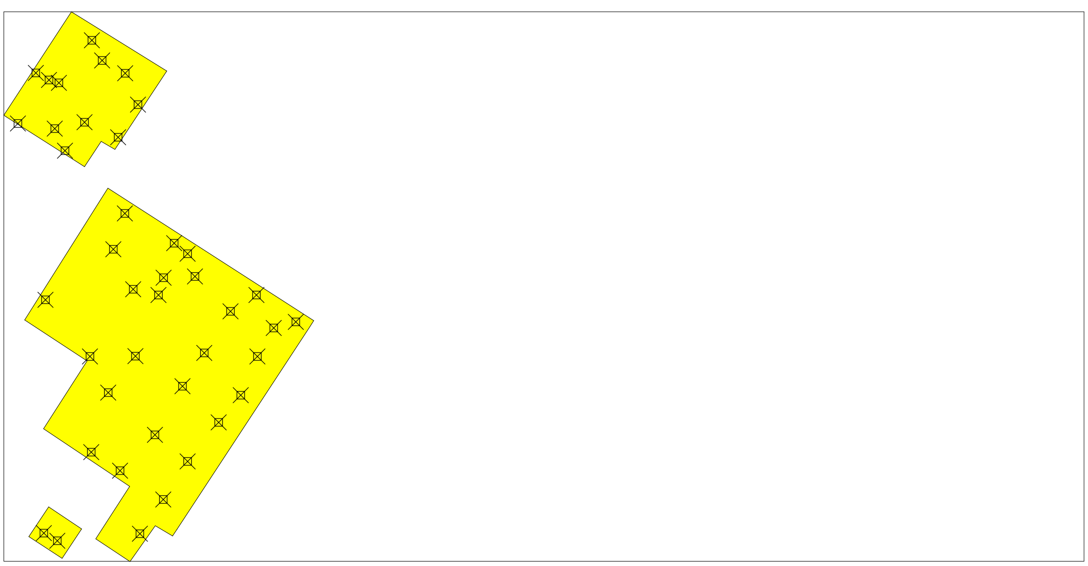
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1

X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.0017	Manual	T
679301.1600	3083254.0340	J-12S	0.59	Manual	T
679171.2970	3083289.7960	J-04S	0.00043	Manual	T
679155.0740	3083294.6960	J-03S	0.0051	Manual	T
679133.4290	3083306.3130	J-01S	0.055	Manual	T
679104.2450	3083223.2620	J-02S	0.019	Manual	T
679164.8060	3083214.7100	J-06S	0.033	Manual	T
679181.2750	3083178.2880	J-08S	0.048	Manual	T
679213.7730	3083224.9730	J-09S	0.00036	Manual	T
679280.5440	3083305.6810	J-10S	0.00038	Manual	T
679242.7260	3083326.5280	J-07S	0.012	Manual	T
679225.8560	3083359.9740	J-05S	0.0021	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	0.013	Manual	T
679335.0020	3082941.1720	J-21S	0.00038	Manual	T
679343.5810	3082969.5980	J-19S	0.0065	Manual	T
679394.8070	3082971.8300	J-24S	0.012	Manual	T
679382.8640	3083009.1130	J-20S	0.00044	Manual	T
679293.5600	3082950.4980	J-17S	0.055	Manual	T
679360.5700	3083026.4980	J-18S	0.0048	Manual	T
679297.0010	3082840.6970	J-23S	0.0077	Manual	T
679252.7130	3082781.0290	J-22S	0.0045	Manual	T
679222.6340	3082840.1720	J-16S	0.0025	Manual	T
679279.6830	3083075.4290	J-14S	0.00038	Manual	T
679149.4920	3082933.0980	J-13S	0.0073	Manual	T
679272.0040	3082652.6750	J-28S	0.054	Manual	T
679224.5850	3082683.1400	J-26S	0.011	Manual	T
679146.6460	3082549.7640	J-25S	0.01	Manual	T
679169.0760	3082537.3510	J-27S	0.0072	Manual	T
679342.7410	3082605.3190	J-35S	0.0038	Manual	T
679304.6530	3082548.6880	J-34S	0.000385	Manual	T
679382.8900	3082667.5270	J-36S	0.008	Manual	T
679329.4380	3082711.0960	J-29S	0.0024	Manual	T
679374.4420	3082791.3300	J-30S	0.0043	Manual	T
679453.4760	3082914.1150	J-32S	0.000365	Manual	T
679410.1490	3082845.8460	J-31S	0.0026	Manual	T
679560.6070	3082897.2580	J-41S	0.0053	Manual	T
679495.8840	3082940.9730	J-33S	0.0048	Manual	T

679524.3310	3082886.8990	J-40S	0.0046	Manual	T
679497.3310	3082840.3960	J-39S	0.014	Manual	T
679470.3570	3082776.7350	J-38S	0.0013	Manual	T
679433.9450	3082731.6820	J-37S	0.00035	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z^2_{1-\alpha}$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

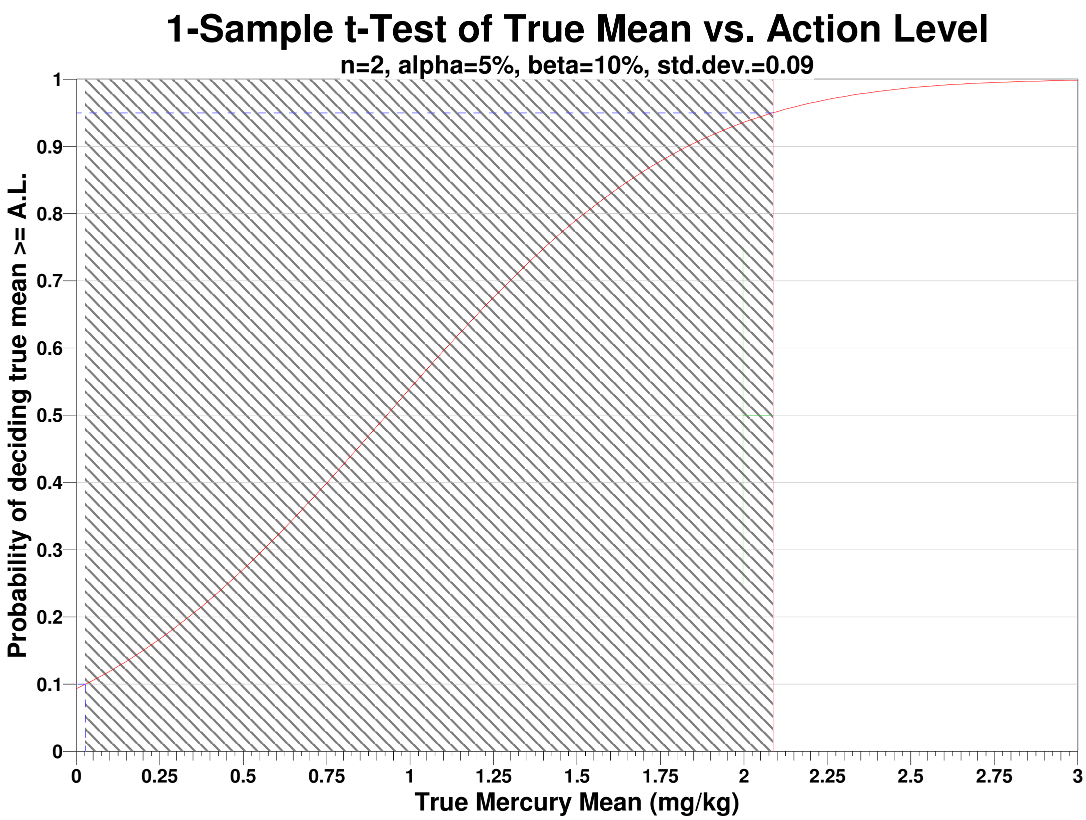
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Mercury	2	0.09 mg/kg	2.06 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples			
AL=2.0872	α=5	α=10	α=15

	s=0.18	s=0.09	s=0.18	s=0.09	s=0.18	s=0.09	
LBGR=90	β=5	10	4	8	3	6	2
	β=10	8	3	6	3	5	2
	β=15	7	3	5	2	4	2
LBGR=80	β=5	4	2	3	2	2	1
	β=10	3	2	3	2	2	1
	β=15	3	2	2	2	2	1
LBGR=70	β=5	3	2	2	1	2	1
	β=10	3	2	2	1	1	1
	β=15	2	2	2	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Mercury
The following data points were entered by the user for analysis.

Mercury (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.00035	0.00036	0.000365	0.00038	0.00038	0.00038	0.000385	0.00043	0.00044	0.0013
10	0.0017	0.0021	0.0024	0.0025	0.0026	0.0038	0.0043	0.0045	0.0046	0.0048
20	0.0048	0.0051	0.0053	0.0065	0.0072	0.0073	0.0077	0.008	0.01	0.011
30	0.012	0.012	0.013	0.014	0.019	0.033	0.048	0.054	0.055	0.055
40	0.59									

SUMMARY STATISTICS for Mercury	
n	41
Min	0.00035

Max				0.59				
Range				0.58965				
Mean				0.02478				
Median				0.0048				
Variance				0.0084248				
StdDev				0.091787				
Std Error				0.014335				
Skewness				6.13				
Interquartile Range				0.0105				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.00035	0.0003605	0.00038	0.0015	0.0048	0.012	0.0528	0.055	0.59

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Mercury			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	6.158	3.05	Yes

The test statistic 6.158 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Mercury	
1	0.59

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.6364
Shapiro-Wilk 5% Critical Value	0.94

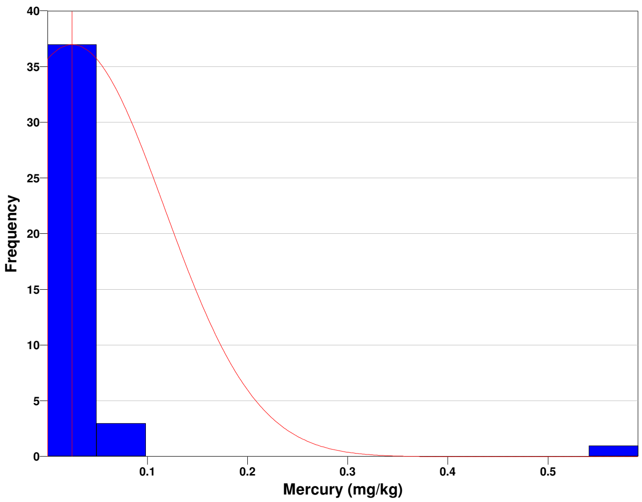
The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

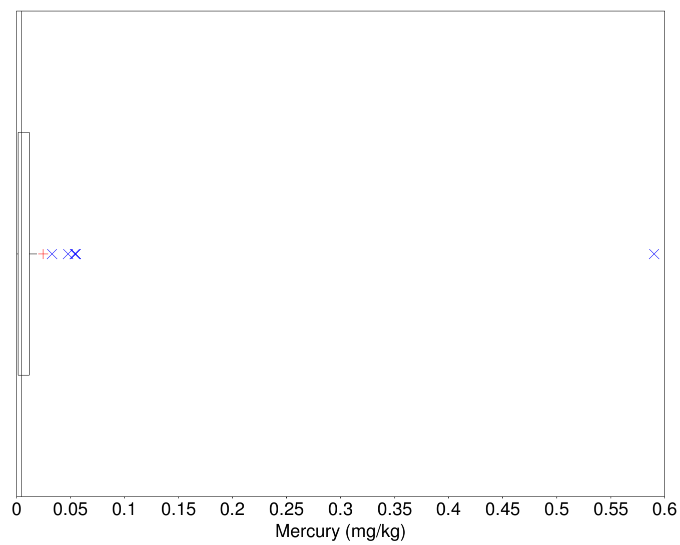
Data Plots for Mercury
Graphical displays of the data are shown below.

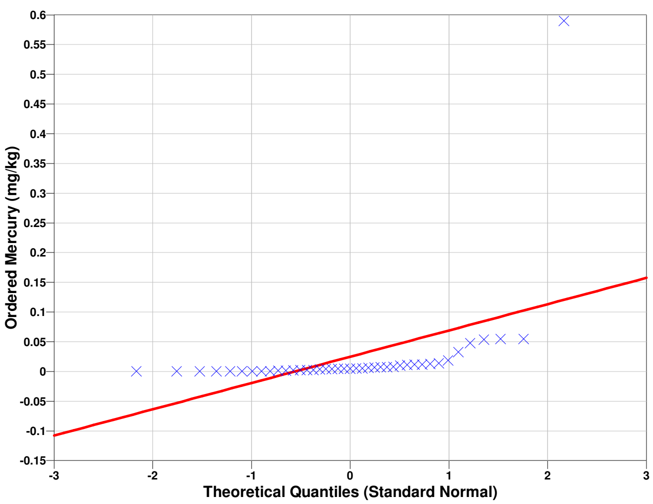
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through

2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Mercury
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.2612
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.04892
95% Non-Parametric (Chebyshev) UCL	0.08726

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.08726) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (2.0872),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-143.88	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

Random sampling locations for comparing a mean with a fixed threshold (parametric)

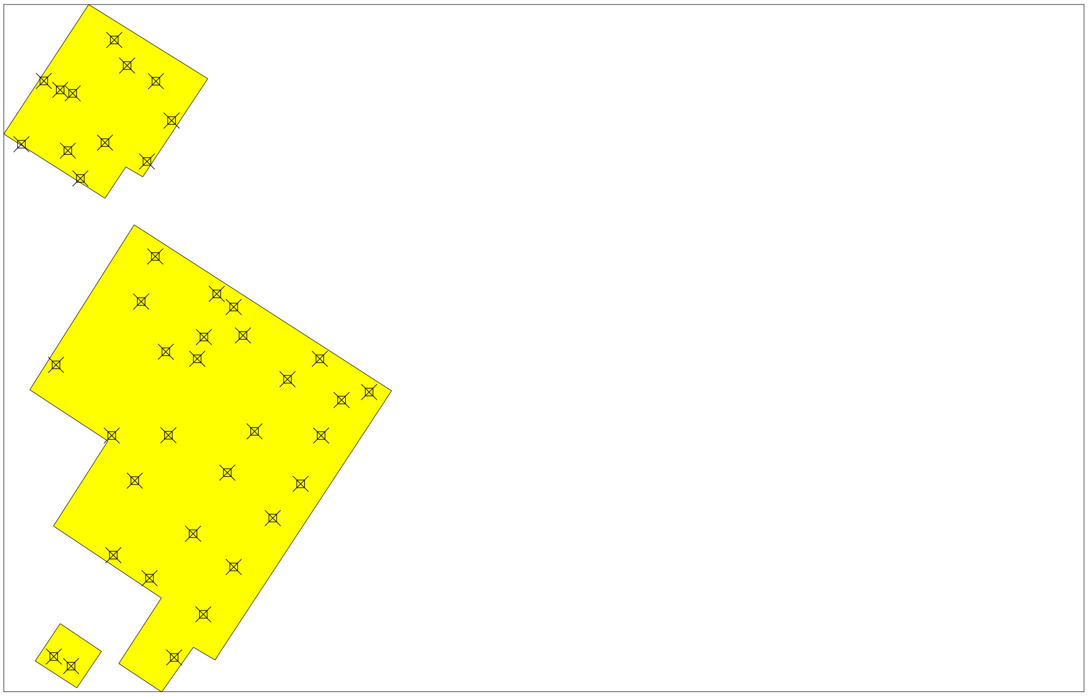
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	2.3	Manual	T
679301.1600	3083254.0340	J-12S	1.3	Manual	T
679171.2970	3083289.7960	J-04S	2.3	Manual	T
679155.0740	3083294.6960	J-03S	2.4	Manual	T
679133.4290	3083306.3130	J-01S	1.9	Manual	T
679104.2450	3083223.2620	J-02S	3.8	Manual	T
679164.8060	3083214.7100	J-06S	3.2	Manual	T
679181.2750	3083178.2880	J-08S	4.1	Manual	T
679213.7730	3083224.9730	J-09S	1.5	Manual	T
679280.5440	3083305.6810	J-10S	1.7	Manual	T
679242.7260	3083326.5280	J-07S	4.3	Manual	T
679225.8560	3083359.9740	J-05S	2.7	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	3.5	Manual	T
679335.0020	3082941.1720	J-21S	1.5	Manual	T
679343.5810	3082969.5980	J-19S	26	Manual	T
679394.8070	3082971.8300	J-24S	3.1	Manual	T
679382.8640	3083009.1130	J-20S	2.1	Manual	T
679293.5600	3082950.4980	J-17S	2.3	Manual	T
679360.5700	3083026.4980	J-18S	4.2	Manual	T

679297.0010	3082840.6970	J-23S	3.6	Manual	T
679252.7130	3082781.0290	J-22S	2.5	Manual	T
679222.6340	3082840.1720	J-16S	2.6	Manual	T
679279.6830	3083075.4290	J-14S	1.8	Manual	T
679149.4920	3082933.0980	J-13S	2.8	Manual	T
679272.0040	3082652.6750	J-28S	9.3	Manual	T
679224.5850	3082683.1400	J-26S	1.9	Manual	T
679146.6460	3082549.7640	J-25S	4.2	Manual	T
679169.0760	3082537.3510	J-27S	3.2	Manual	T
679342.7410	3082605.3190	J-35S	2	Manual	T
679304.6530	3082548.6880	J-34S	1.9	Manual	T
679382.8900	3082667.5270	J-36S	6.8	Manual	T
679329.4380	3082711.0960	J-29S	2.3	Manual	T
679374.4420	3082791.3300	J-30S	2.7	Manual	T
679453.4760	3082914.1150	J-32S	2.7	Manual	T
679410.1490	3082845.8460	J-31S	1.7	Manual	T
679560.6070	3082897.2580	J-41S	6.3	Manual	T
679495.8840	3082940.9730	J-33S	5.9	Manual	T
679524.3310	3082886.8990	J-40S	2.15	Manual	T
679497.3310	3082840.3960	J-39S	1.6	Manual	T
679470.3570	3082776.7350	J-38S	5.7	Manual	T
679433.9450	3082731.6820	J-37S	6	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability (1-β) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

- n is the number of samples,
- S is the estimated standard deviation of the measured values including analytical error,
- Δ is the width of the gray region,
- α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
- β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
- $Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is 1-α,
- $Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is 1-β.

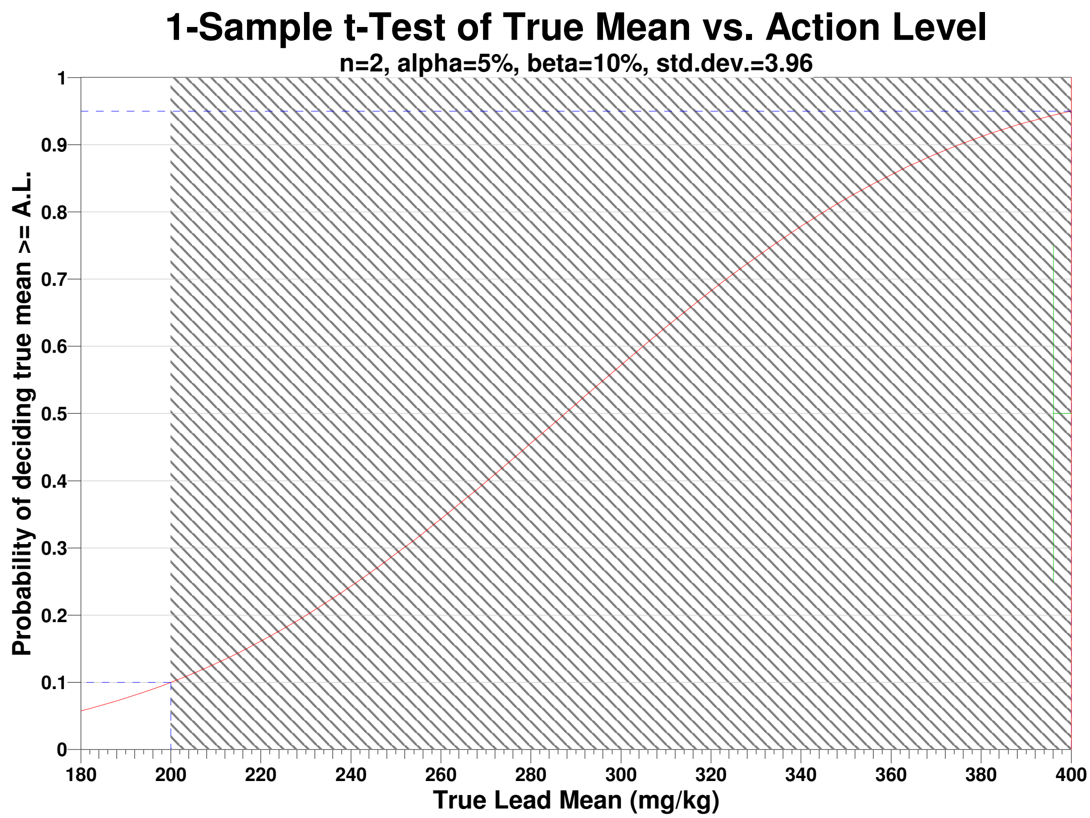
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	Δ	α	β	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Lead	2	3.96 mg/kg	200 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at 1-α on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at 1-α. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=400		α=5		α=10		α=15	
		s=7.92	s=3.96	s=7.92	s=3.96	s=7.92	s=3.96
LBGR=90	β=5	2	2	2	1	1	1
	β=10	2	2	2	1	1	1
	β=15	2	2	2	1	1	1
LBGR=80	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	β=5	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Lead
The following data points were entered by the user for analysis.

Lead (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	1.3	1.5	1.5	1.6	1.7	1.7	1.8	1.9	1.9	1.9
10	2	2.1	2.15	2.3	2.3	2.3	2.3	2.4	2.5	2.6
20	2.7	2.7	2.7	2.8	3.1	3.2	3.2	3.5	3.6	3.8
30	4.1	4.2	4.2	4.3	5.7	5.9	6	6.3	6.8	9.3
40	26									

SUMMARY STATISTICS for Lead	
n	41
Min	1.3
Max	26
Range	24.7
Mean	3.7524
Median	2.7
Variance	15.65
StdDev	3.956
Std Error	0.61782
Skewness	4.7237

Interquartile Range				2.2				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
1.3	1.5	1.62	1.95	2.7	4.15	6.24	9.05	26

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Lead			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	5.624	3.05	Yes

The test statistic 5.624 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Lead	
1	26

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.835
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

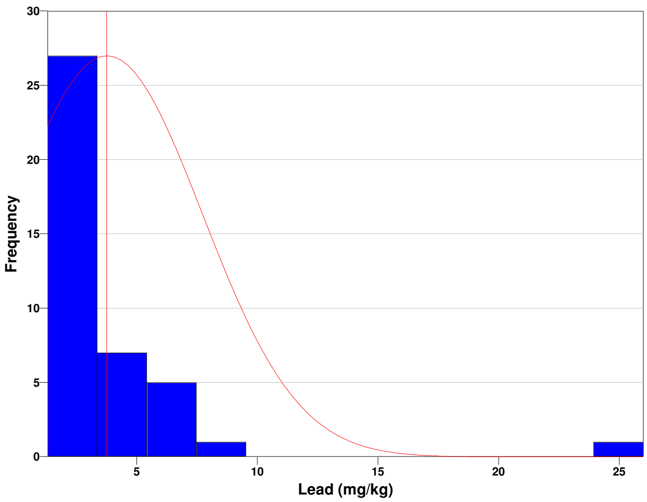
Data Plots for Lead
Graphical displays of the data are shown below.

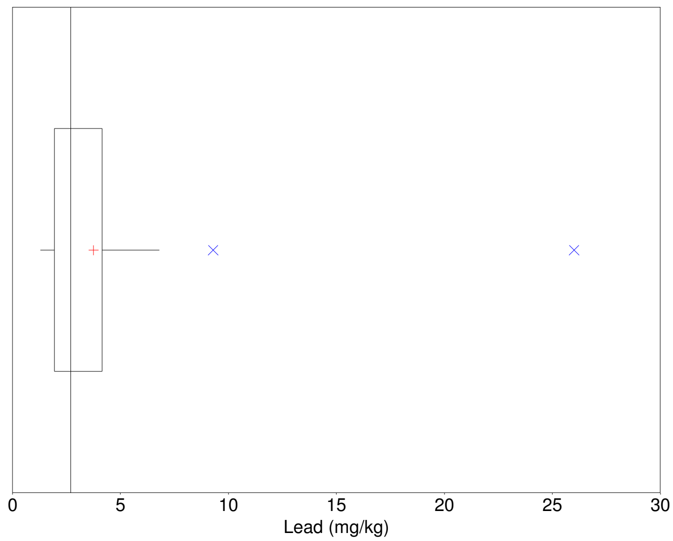
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

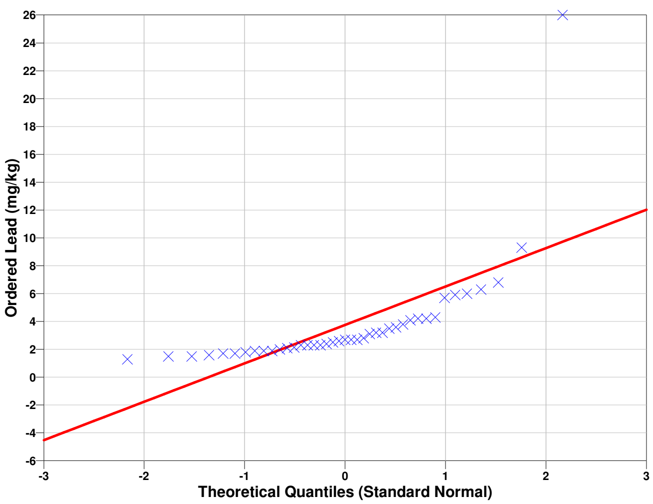
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Lead
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.5047
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	4.793
95% Non-Parametric (Chebyshev) UCL	6.445

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (6.445) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (400),
SE is the standard error = (standard deviation) / (square root of n).

This *t* was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-641.36	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

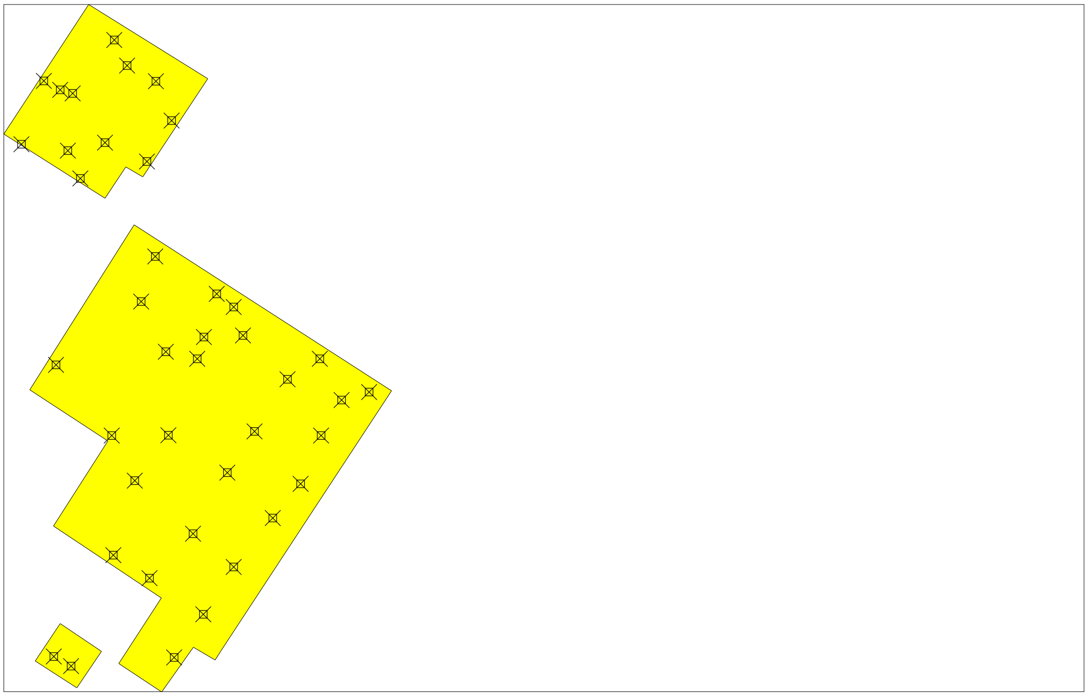
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	2.3	Manual	T
679301.1600	3083254.0340	J-12S	1.3	Manual	T
679171.2970	3083289.7960	J-04S	2.3	Manual	T
679155.0740	3083294.6960	J-03S	2.4	Manual	T
679133.4290	3083306.3130	J-01S	1.9	Manual	T
679104.2450	3083223.2620	J-02S	3.8	Manual	T
679164.8060	3083214.7100	J-06S	3.2	Manual	T
679181.2750	3083178.2880	J-08S	4.1	Manual	T
679213.7730	3083224.9730	J-09S	1.5	Manual	T
679280.5440	3083305.6810	J-10S	1.7	Manual	T
679242.7260	3083326.5280	J-07S	4.3	Manual	T
679225.8560	3083359.9740	J-05S	2.7	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	3.5	Manual	T
679335.0020	3082941.1720	J-21S	1.5	Manual	T
679343.5810	3082969.5980	J-19S	26	Manual	T
679394.8070	3082971.8300	J-24S	3.1	Manual	T
679382.8640	3083009.1130	J-20S	2.1	Manual	T
679293.5600	3082950.4980	J-17S	2.3	Manual	T
679360.5700	3083026.4980	J-18S	4.2	Manual	T

679297.0010	3082840.6970	J-23S	3.6	Manual	T
679252.7130	3082781.0290	J-22S	2.5	Manual	T
679222.6340	3082840.1720	J-16S	2.6	Manual	T
679279.6830	3083075.4290	J-14S	1.8	Manual	T
679149.4920	3082933.0980	J-13S	2.8	Manual	T
679272.0040	3082652.6750	J-28S	9.3	Manual	T
679224.5850	3082683.1400	J-26S	1.9	Manual	T
679146.6460	3082549.7640	J-25S	4.2	Manual	T
679169.0760	3082537.3510	J-27S	3.2	Manual	T
679342.7410	3082605.3190	J-35S	2	Manual	T
679304.6530	3082548.6880	J-34S	1.9	Manual	T
679382.8900	3082667.5270	J-36S	6.8	Manual	T
679329.4380	3082711.0960	J-29S	2.3	Manual	T
679374.4420	3082791.3300	J-30S	2.7	Manual	T
679453.4760	3082914.1150	J-32S	2.7	Manual	T
679410.1490	3082845.8460	J-31S	1.7	Manual	T
679560.6070	3082897.2580	J-41S	6.3	Manual	T
679495.8840	3082940.9730	J-33S	5.9	Manual	T
679524.3310	3082886.8990	J-40S	2.15	Manual	T
679497.3310	3082840.3960	J-39S	1.6	Manual	T
679470.3570	3082776.7350	J-38S	5.7	Manual	T
679433.9450	3082731.6820	J-37S	6	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

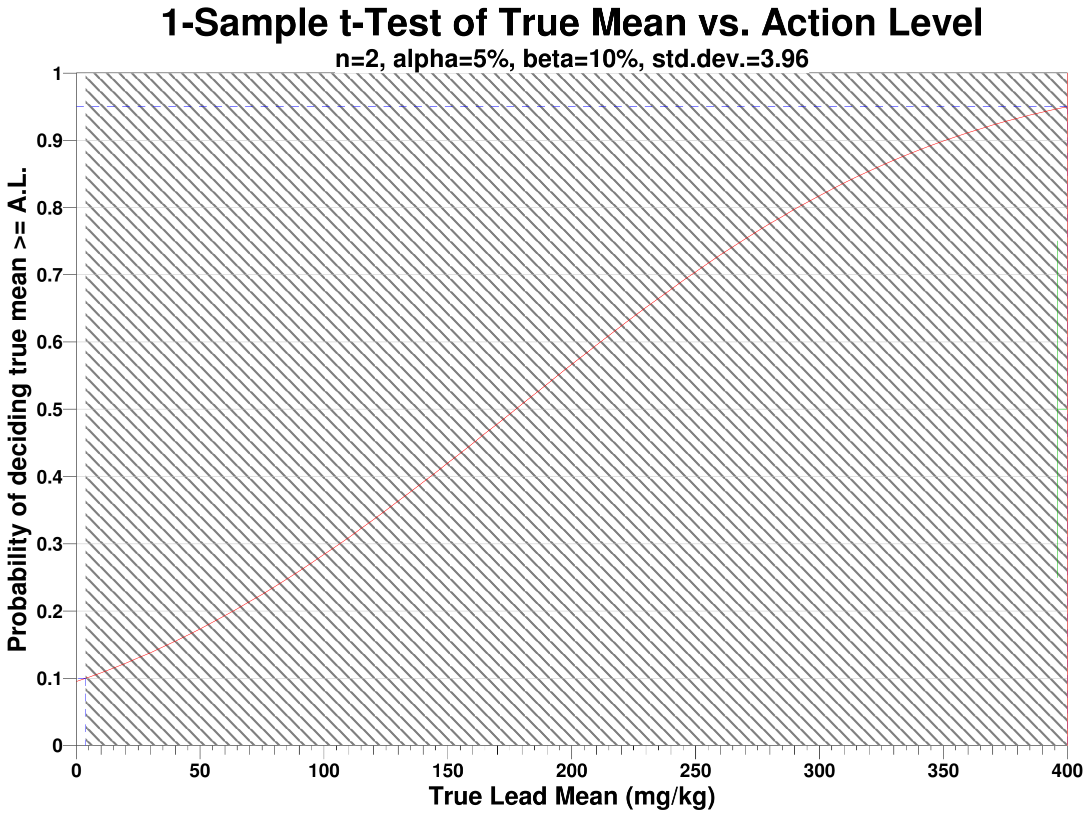
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Lead	2	3.96 mg/kg	396.25 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=400		α=5		α=10		α=15	
		s=7.92	s=3.96	s=7.92	s=3.96	s=7.92	s=3.96
LBGR=90	β=5	2	2	2	1	1	1
	β=10	2	2	2	1	1	1
	β=15	2	2	2	1	1	1
LBGR=80	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	β=5	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Lead
The following data points were entered by the user for analysis.

Lead (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	1.3	1.5	1.5	1.6	1.7	1.7	1.8	1.9	1.9	1.9
10	2	2.1	2.15	2.3	2.3	2.3	2.3	2.4	2.5	2.6
20	2.7	2.7	2.7	2.8	3.1	3.2	3.2	3.5	3.6	3.8
30	4.1	4.2	4.2	4.3	5.7	5.9	6	6.3	6.8	9.3
40	26									

SUMMARY STATISTICS for Lead	
n	41
Min	1.3
Max	26
Range	24.7
Mean	3.7524
Median	2.7
Variance	15.65
StdDev	3.956
Std Error	0.61782
Skewness	4.7237

Interquartile Range				2.2				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
1.3	1.5	1.62	1.95	2.7	4.15	6.24	9.05	26

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Lead			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	5.624	3.05	Yes

The test statistic 5.624 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Lead	
1	26

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.835
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

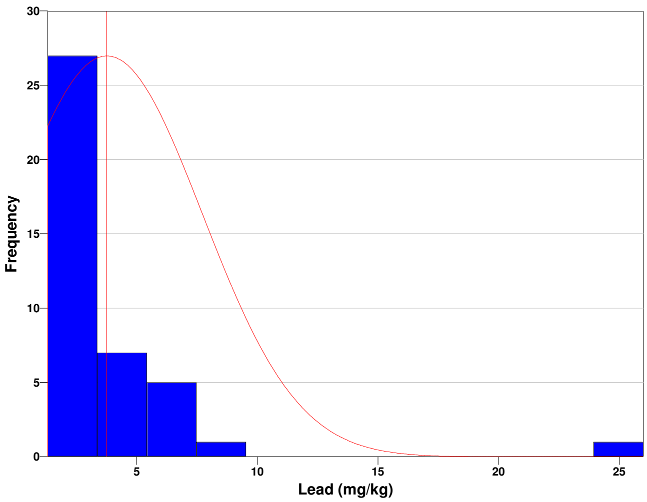
Data Plots for Lead
Graphical displays of the data are shown below.

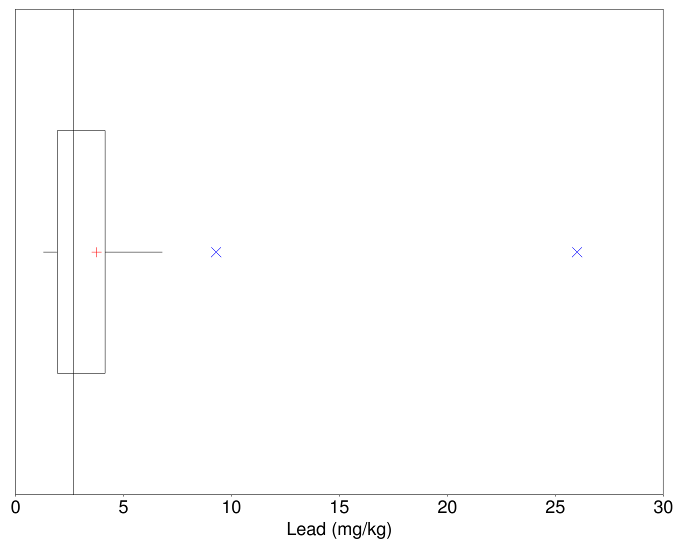
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

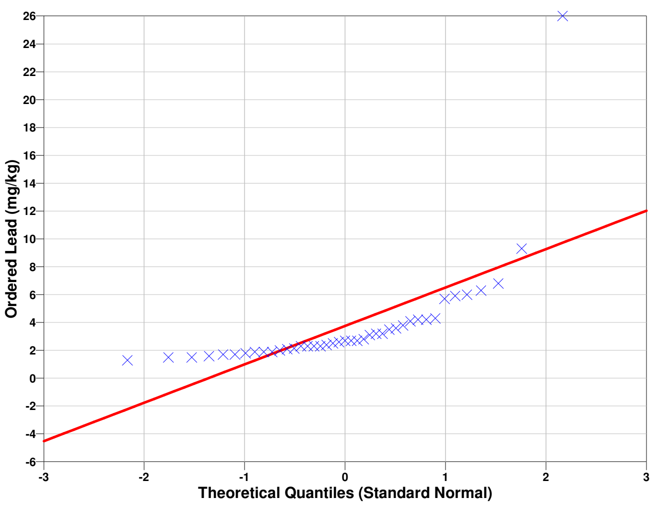
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Lead

A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.5047
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	4.793
95% Non-Parametric (Chebyshev) UCL	6.445

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (6.445) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (400),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-641.36	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

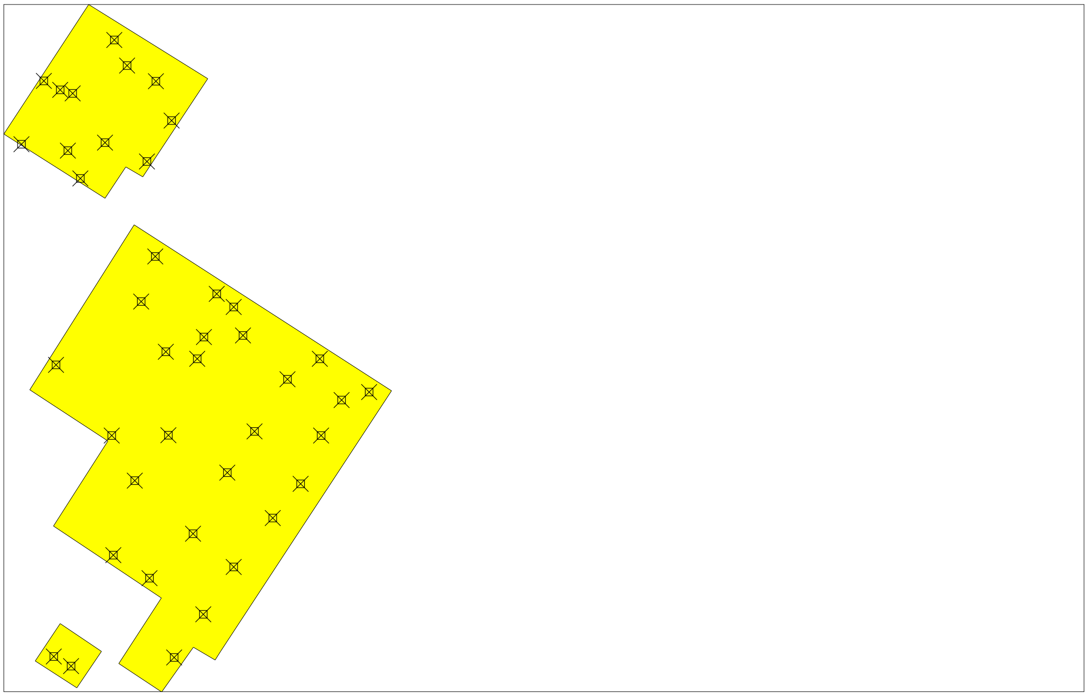
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.89	Manual	T
679301.1600	3083254.0340	J-12S	0.65	Manual	T
679171.2970	3083289.7960	J-04S	1	Manual	T
679155.0740	3083294.6960	J-03S	1.7	Manual	T
679133.4290	3083306.3130	J-01S	0.97	Manual	T
679104.2450	3083223.2620	J-02S	5	Manual	T
679164.8060	3083214.7100	J-06S	5.3	Manual	T
679181.2750	3083178.2880	J-08S	5.7	Manual	T
679213.7730	3083224.9730	J-09S	0.95	Manual	T
679280.5440	3083305.6810	J-10S	0.83	Manual	T
679242.7260	3083326.5280	J-07S	7.1	Manual	T
679225.8560	3083359.9740	J-05S	0.77	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	4.4	Manual	T
679335.0020	3082941.1720	J-21S	1.1	Manual	T
679343.5810	3082969.5980	J-19S	3.2	Manual	T
679394.8070	3082971.8300	J-24S	1.2	Manual	T
679382.8640	3083009.1130	J-20S	1.8	Manual	T
679293.5600	3082950.4980	J-17S	2.1	Manual	T
679360.5700	3083026.4980	J-18S	3.6	Manual	T

679297.0010	3082840.6970	J-23S	6.7	Manual	T
679252.7130	3082781.0290	J-22S	2.3	Manual	T
679222.6340	3082840.1720	J-16S	4.5	Manual	T
679279.6830	3083075.4290	J-14S	1.3	Manual	T
679149.4920	3082933.0980	J-13S	6.9	Manual	T
679272.0040	3082652.6750	J-28S	13	Manual	T
679224.5850	3082683.1400	J-26S	2.9	Manual	T
679146.6460	3082549.7640	J-25S	13.7	Manual	T
679169.0760	3082537.3510	J-27S	6	Manual	T
679342.7410	3082605.3190	J-35S	3.1	Manual	T
679304.6530	3082548.6880	J-34S	2.4	Manual	T
679382.8900	3082667.5270	J-36S	11	Manual	T
679329.4380	3082711.0960	J-29S	3.8	Manual	T
679374.4420	3082791.3300	J-30S	2.4	Manual	T
679453.4760	3082914.1150	J-32S	3.7	Manual	T
679410.1490	3082845.8460	J-31S	1.7	Manual	T
679560.6070	3082897.2580	J-41S	4.3	Manual	T
679495.8840	3082940.9730	J-33S	2	Manual	T
679524.3310	3082886.8990	J-40S	1.95	Manual	T
679497.3310	3082840.3960	J-39S	2.4	Manual	T
679470.3570	3082776.7350	J-38S	5.8	Manual	T
679433.9450	3082731.6820	J-37S	12.1	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

The values of these inputs that result in the calculated number of sampling locations are:

Analyte	**n**	**Parameter**					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Vanadium	2	3.4 mg/kg	145.507 mg/kg	0.05	0.1	1.64485	1.28155

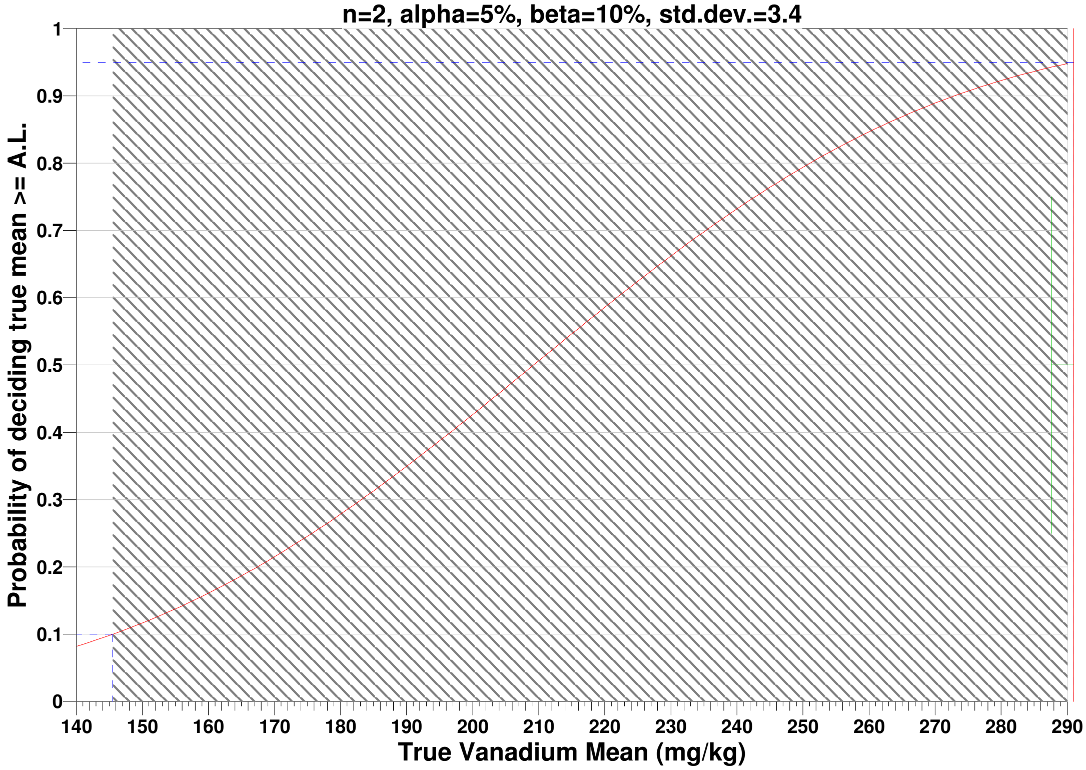
[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.

1-Sample t-Test of True Mean vs. Action Level

n=2, alpha=5%, beta=10%, std.dev.=3.4



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=291.014		α=5		α=10		α=15	
		s=6.8	s=3.4	s=6.8	s=3.4	s=6.8	s=3.4
LBGR=90	β=5	2	2	2	1	1	1
	β=10	2	2	2	1	1	1
	β=15	2	2	2	1	1	1
LBGR=80	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	β=5	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Vanadium
The following data points were entered by the user for analysis.

Vanadium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.65	0.77	0.83	0.89	0.95	0.97	1	1.1	1.2	1.3
10	1.7	1.7	1.8	1.95	2	2.1	2.3	2.4	2.4	2.4
20	2.9	3.1	3.2	3.6	3.7	3.8	4.3	4.4	4.5	5
30	5.3	5.7	5.8	6	6.7	6.9	7.1	11	12.1	13
40	13.7									

SUMMARY STATISTICS for Vanadium	
n	41
Min	0.65
Max	13.7
Range	13.05
Mean	3.9563
Median	2.9
Variance	11.535
StdDev	3.3963
Std Error	0.53042
Skewness	1.5509

Interquartile Range				4				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.65	0.776	0.902	1.5	2.9	5.5	10.22	12.91	13.7

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Vanadium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	2.869	3.05	No

None of the test statistics exceeded the corresponding critical values, therefore none of the 1 tests are significant and we conclude that at the 5% significance level there are no outliers in the data.

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.8286
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

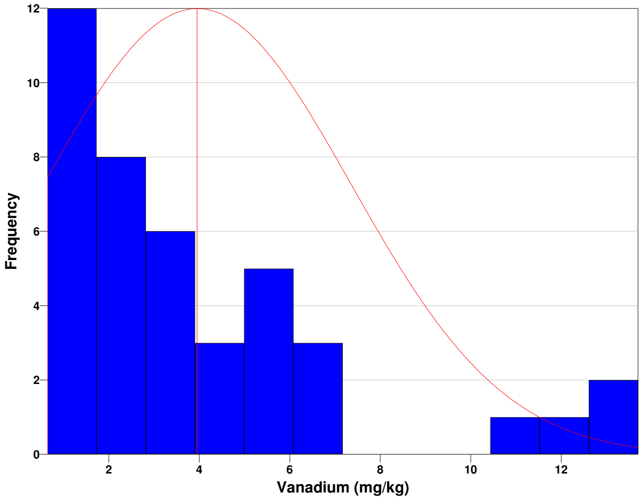
Data Plots for Vanadium
Graphical displays of the data are shown below.

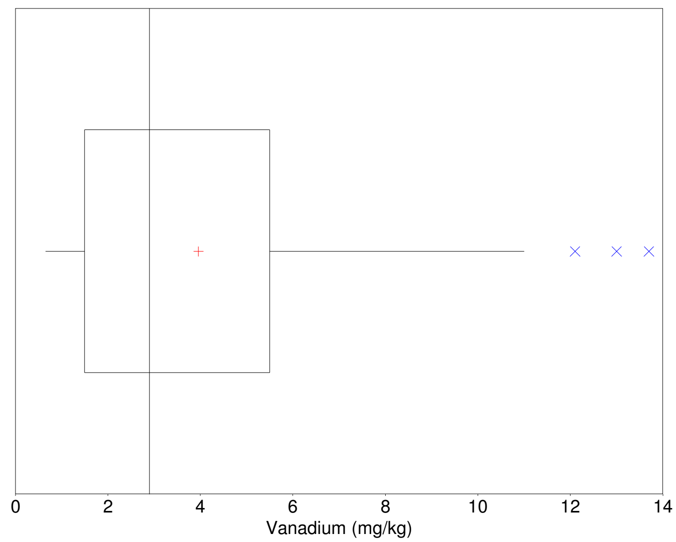
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

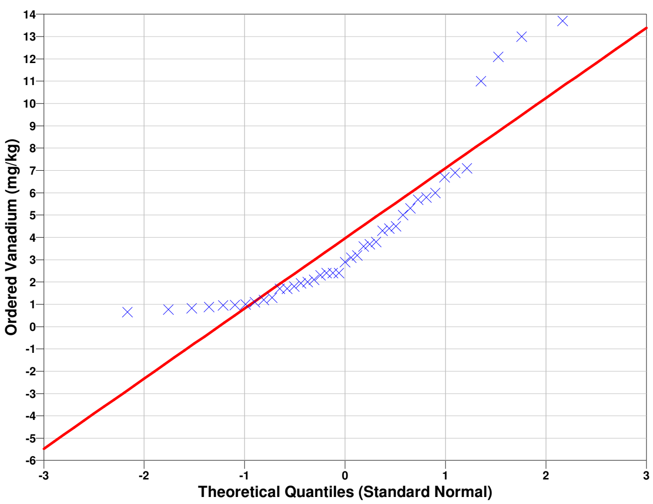
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the

distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Vanadium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.8131
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	4.849
95% Non-Parametric (Chebyshev) UCL	6.268

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (6.268) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (291.014),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-541.19	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

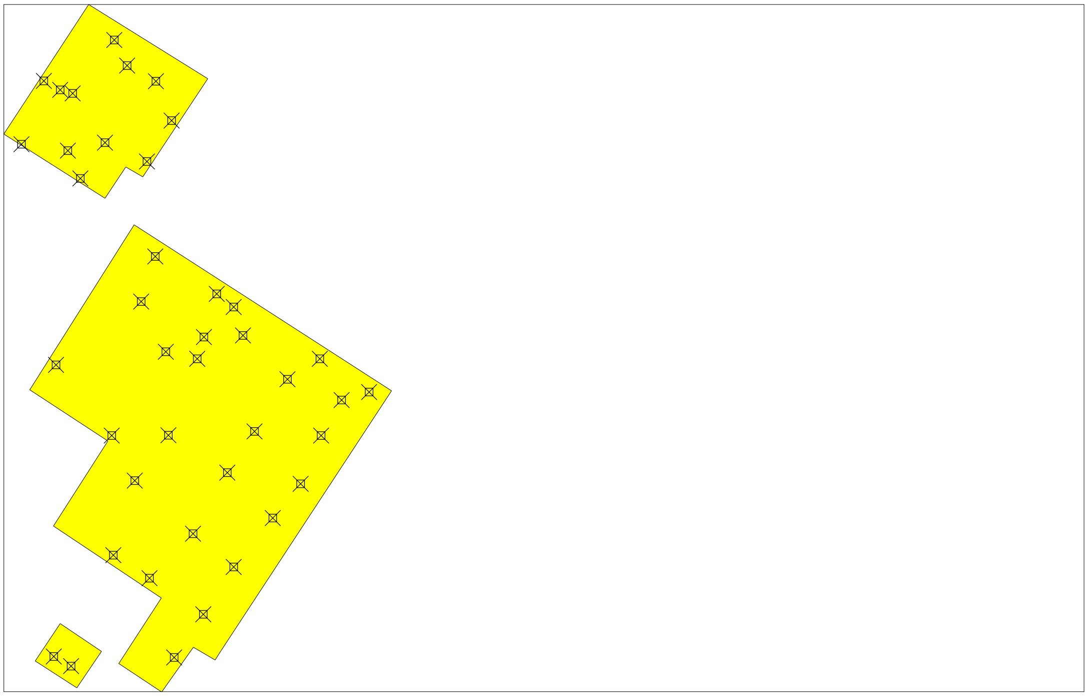
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.89	Manual	T
679301.1600	3083254.0340	J-12S	0.65	Manual	T
679171.2970	3083289.7960	J-04S	1	Manual	T
679155.0740	3083294.6960	J-03S	1.7	Manual	T
679133.4290	3083306.3130	J-01S	0.97	Manual	T
679104.2450	3083223.2620	J-02S	5	Manual	T
679164.8060	3083214.7100	J-06S	5.3	Manual	T
679181.2750	3083178.2880	J-08S	5.7	Manual	T
679213.7730	3083224.9730	J-09S	0.95	Manual	T
679280.5440	3083305.6810	J-10S	0.83	Manual	T
679242.7260	3083326.5280	J-07S	7.1	Manual	T
679225.8560	3083359.9740	J-05S	0.77	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	4.4	Manual	T
679335.0020	3082941.1720	J-21S	1.1	Manual	T
679343.5810	3082969.5980	J-19S	3.2	Manual	T
679394.8070	3082971.8300	J-24S	1.2	Manual	T
679382.8640	3083009.1130	J-20S	1.8	Manual	T
679293.5600	3082950.4980	J-17S	2.1	Manual	T
679360.5700	3083026.4980	J-18S	3.6	Manual	T

679297.0010	3082840.6970	J-23S	6.7	Manual	T
679252.7130	3082781.0290	J-22S	2.3	Manual	T
679222.6340	3082840.1720	J-16S	4.5	Manual	T
679279.6830	3083075.4290	J-14S	1.3	Manual	T
679149.4920	3082933.0980	J-13S	6.9	Manual	T
679272.0040	3082652.6750	J-28S	13	Manual	T
679224.5850	3082683.1400	J-26S	2.9	Manual	T
679146.6460	3082549.7640	J-25S	13.7	Manual	T
679169.0760	3082537.3510	J-27S	6	Manual	T
679342.7410	3082605.3190	J-35S	3.1	Manual	T
679304.6530	3082548.6880	J-34S	2.4	Manual	T
679382.8900	3082667.5270	J-36S	11	Manual	T
679329.4380	3082711.0960	J-29S	3.8	Manual	T
679374.4420	3082791.3300	J-30S	2.4	Manual	T
679453.4760	3082914.1150	J-32S	3.7	Manual	T
679410.1490	3082845.8460	J-31S	1.7	Manual	T
679560.6070	3082897.2580	J-41S	4.3	Manual	T
679495.8840	3082940.9730	J-33S	2	Manual	T
679524.3310	3082886.8990	J-40S	1.95	Manual	T
679497.3310	3082840.3960	J-39S	2.4	Manual	T
679470.3570	3082776.7350	J-38S	5.8	Manual	T
679433.9450	3082731.6820	J-37S	12.1	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

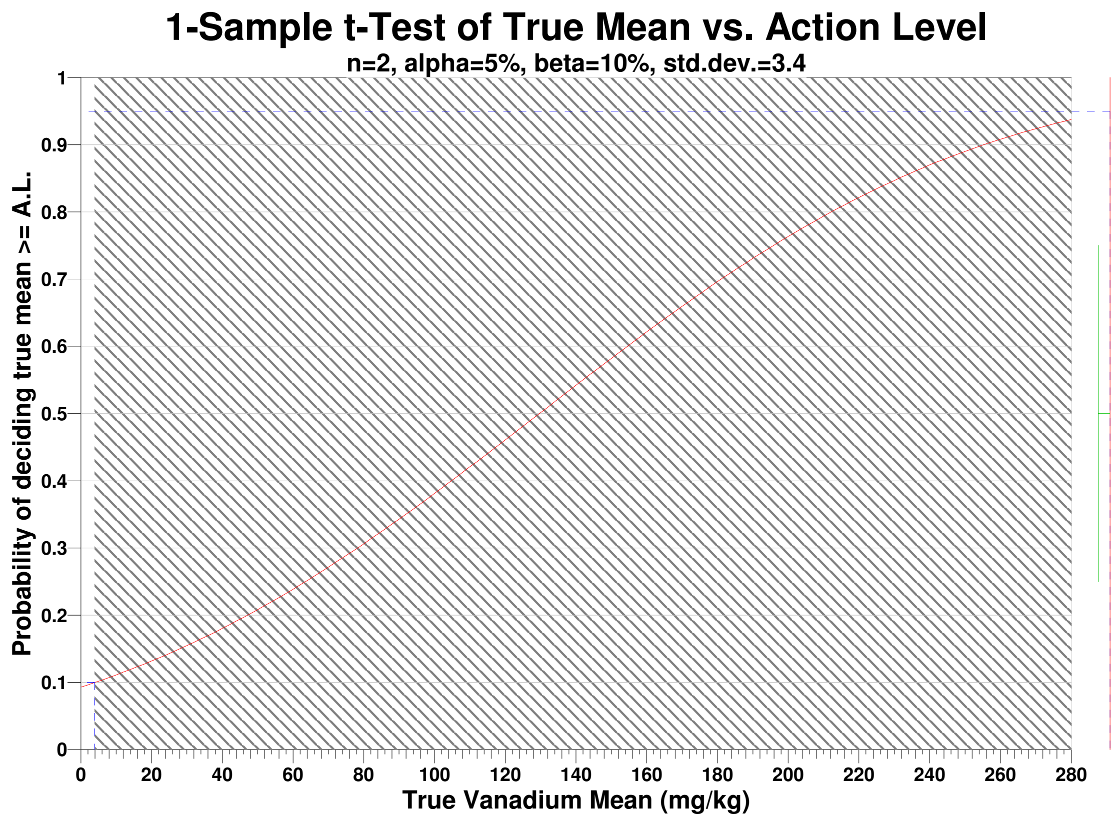
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	**n**	**Parameter**					
		S	**Δ**	**α**	**β**	***Z*** $Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Vanadium	2	3.4 mg/kg	287.06 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=291.014		α=5		α=10		α=15	
		s=6.8	s=3.4	s=6.8	s=3.4	s=6.8	s=3.4
LBGR=90	β=5	2	2	2	1	1	1
	β=10	2	2	2	1	1	1
	β=15	2	2	2	1	1	1
LBGR=80	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	β=5	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Vanadium
The following data points were entered by the user for analysis.

Vanadium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.65	0.77	0.83	0.89	0.95	0.97	1	1.1	1.2	1.3
10	1.7	1.7	1.8	1.95	2	2.1	2.3	2.4	2.4	2.4
20	2.9	3.1	3.2	3.6	3.7	3.8	4.3	4.4	4.5	5
30	5.3	5.7	5.8	6	6.7	6.9	7.1	11	12.1	13
40	13.7									

SUMMARY STATISTICS for Vanadium	
n	41
Min	0.65
Max	13.7
Range	13.05
Mean	3.9563
Median	2.9
Variance	11.535
StdDev	3.3963
Std Error	0.53042
Skewness	1.5509

Interquartile Range				4				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.65	0.776	0.902	1.5	2.9	5.5	10.22	12.91	13.7

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Vanadium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	2.869	3.05	No

None of the test statistics exceeded the corresponding critical values, therefore none of the 1 tests are significant and we conclude that at the 5% significance level there are no outliers in the data.

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.8286
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

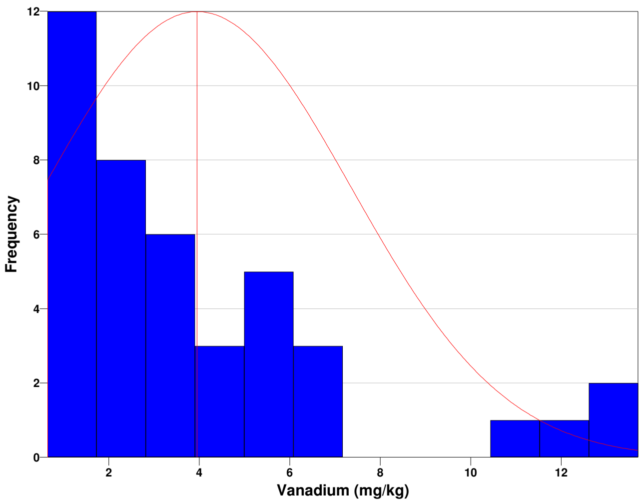
Data Plots for Vanadium
Graphical displays of the data are shown below.

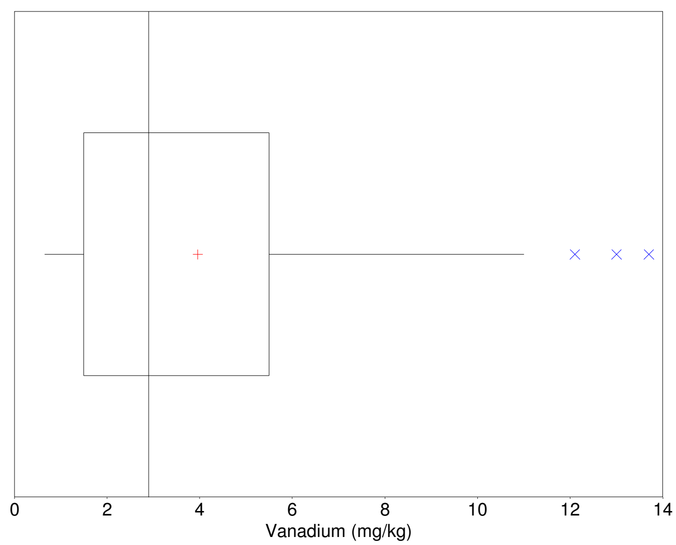
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

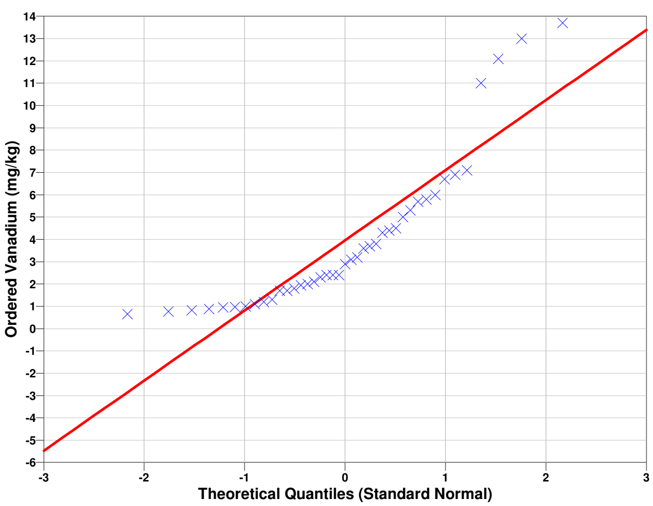
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the

distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Vanadium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.8131
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	4.849
95% Non-Parametric (Chebyshev) UCL	6.268

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (6.268) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (291.014),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-541.19	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

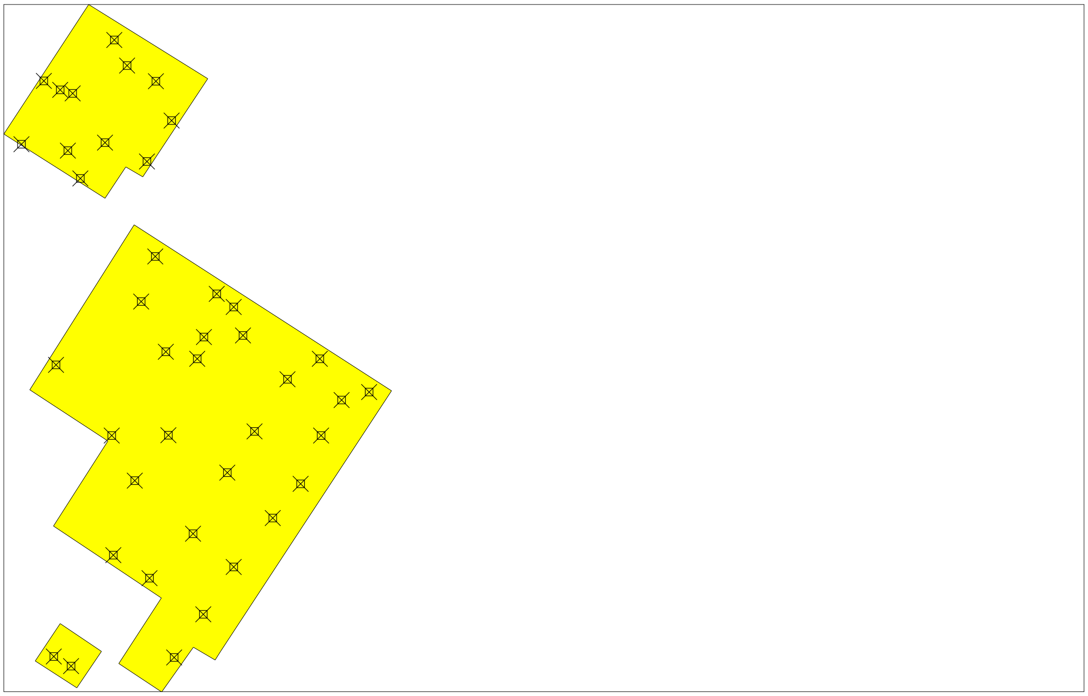
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	23
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$12,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679133.4290	3083306.3130	J-01S	3010	Manual	T
679104.2450	3083223.2620	J-02S	820	Manual	T
679155.0740	3083294.6960	J-03S	659.5	Manual	T
679171.2970	3083289.7960	J-04S	786	Manual	T
679225.8560	3083359.9740	J-05S	1250	Manual	T
679164.8060	3083214.7100	J-06S	609	Manual	T
679242.7260	3083326.5280	J-07S	2900	Manual	T
679181.2750	3083178.2880	J-08S	5020	Manual	T
679213.7730	3083224.9730	J-09S	5460	Manual	T
679280.5440	3083305.6810	J-10S	1670	Manual	T
679268.7700	3083200.3260	J-11S	14300	Manual	T
679301.1600	3083254.0340	J-12S	3550	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679149.4920	3082933.0980	J-13S	2590	Manual	T
679279.6830	3083075.4290	J-14S	2710	Manual	T
679261.0980	3083016.3510	J-15S	1440	Manual	T
679222.6340	3082840.1720	J-16S	3680	Manual	T
679293.5600	3082950.4980	J-17S	7070	Manual	T
679360.5700	3083026.4980	J-18S	3030	Manual	T
679343.5810	3082969.5980	J-19S	4375	Manual	T

679382.8640	3083009.1130	J-20S	4570	Manual	T
679335.0020	3082941.1720	J-21S	4470	Manual	T
679252.7130	3082781.0290	J-22S	5620	Manual	T
679297.0010	3082840.6970	J-23S	8090	Manual	T
679394.8070	3082971.8300	J-24S	2110	Manual	T
679146.6460	3082549.7640	J-25S	1105	Manual	T
679224.5850	3082683.1400	J-26S	2540.05	Manual	T
679169.0760	3082537.3510	J-27S	2550	Manual	T
679272.0040	3082652.6750	J-28S	5510	Manual	T
679329.4380	3082711.0960	J-29S	12200	Manual	T
679374.4420	3082791.3300	J-30S	14400	Manual	T
679410.1490	3082845.8460	J-31S	15650	Manual	T
679453.4760	3082914.1150	J-32S	14200	Manual	T
679495.8840	3082940.9730	J-33S	2600	Manual	T
679304.6530	3082548.6880	J-34S	25400	Manual	T
679342.7410	3082605.3190	J-35S	5130	Manual	T
679382.8900	3082667.5270	J-36S	7830	Manual	T
679433.9450	3082731.6820	J-37S	6190	Manual	T
679470.3570	3082776.7350	J-38S	9900	Manual	T
679497.3310	3082840.3960	J-39S	5110	Manual	T
679524.3310	3082886.8990	J-40S	9850	Manual	T
679560.6070	3082897.2580	J-41S	4830	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

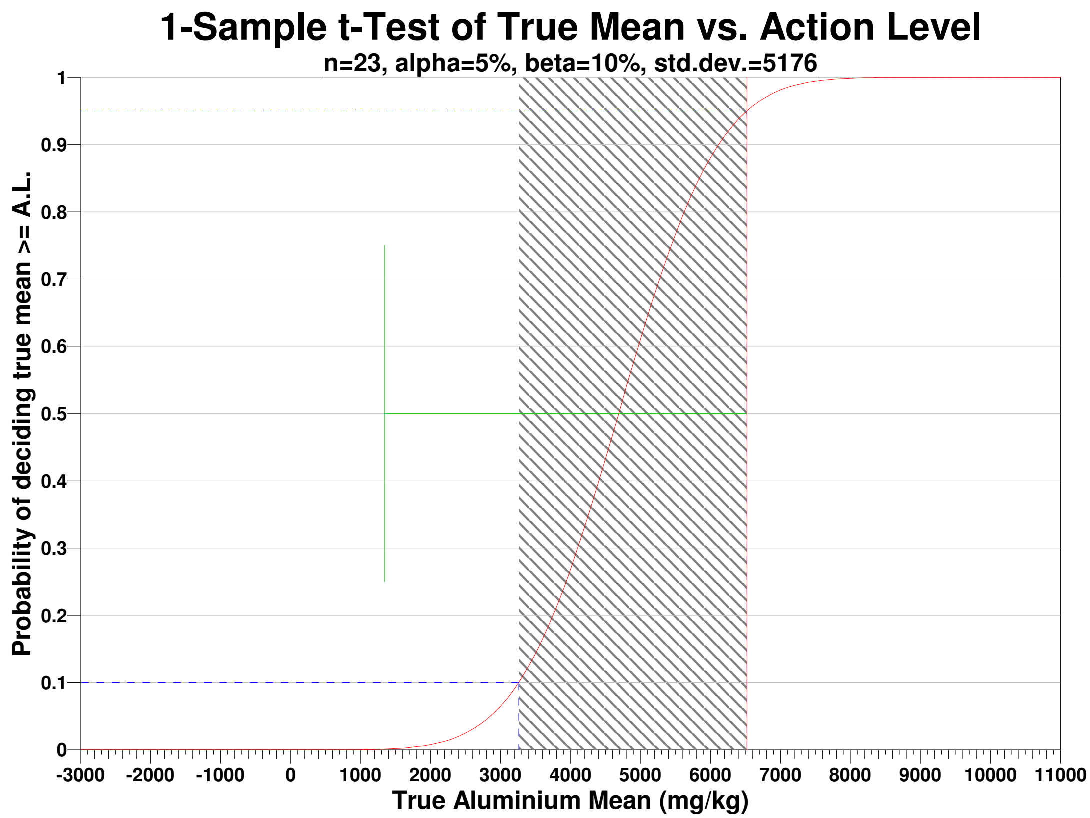
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	**n**	**Parameter**					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Aluminium	23	5176 mg/kg	3261 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=6521		α=5		α=10		α=15	
		s=10352	s=5176	s=10352	s=5176	s=10352	s=5176
LBGR=90	β=5	2729	684	2160	541	1813	454
	β=10	2160	541	1657	415	1355	340
	β=15	1814	455	1355	340	1084	272
LBGR=80	β=5	684	172	541	136	454	114
	β=10	541	137	415	105	340	86
	β=15	455	115	340	86	272	69
LBGR=70	β=5	305	78	241	61	202	51

	β=10	242	62	185	47	151	39
	β=15	203	52	152	39	121	31

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $12,500.00, which averages out to a per sample cost of $543.48. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**23 Samples**
Field collection costs		$100.00	$2,300.00
Analytical costs	$400.00	$400.00	$9,200.00
Sum of Field & Analytical costs		**$500.00**	**$11,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$12,500.00**

Data Analysis for Aluminium
The following data points were entered by the user for analysis.

Aluminium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	609	659.5	786	820	1105	1250	1440	1670	2110	2540
10	2550	2590	2600	2710	2900	3010	3030	3550	3680	4375
20	4470	4570	4830	5020	5110	5130	5460	5510	5620	6190
30	7070	7830	8090	9850	9900	1.22e+004	1.42e+004	1.43e+004	1.44e+004	1.565e+004
40	2.54e+004									

SUMMARY STATISTICS for Aluminium	
n	41
Min	609
Max	25400
Range	24791
Mean	5726.5
Median	4470
Variance	2.6787e+007
StdDev	5175.7
Std Error	808.3
Skewness	1.857

Interquartile Range				4905				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
609	672.1	877	2545	4470	7450	1.428e+004	1.552e+004	2.54e+004

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Aluminium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.801	3.05	Yes

The test statistic 3.801 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Aluminium	
1	25400

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.8538
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

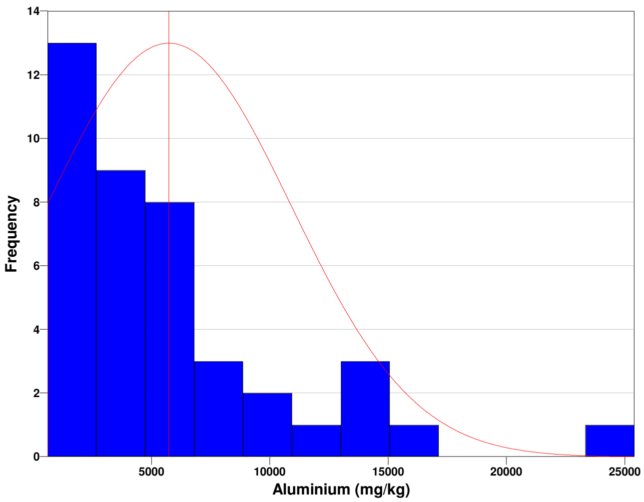
Data Plots for Aluminium
Graphical displays of the data are shown below.

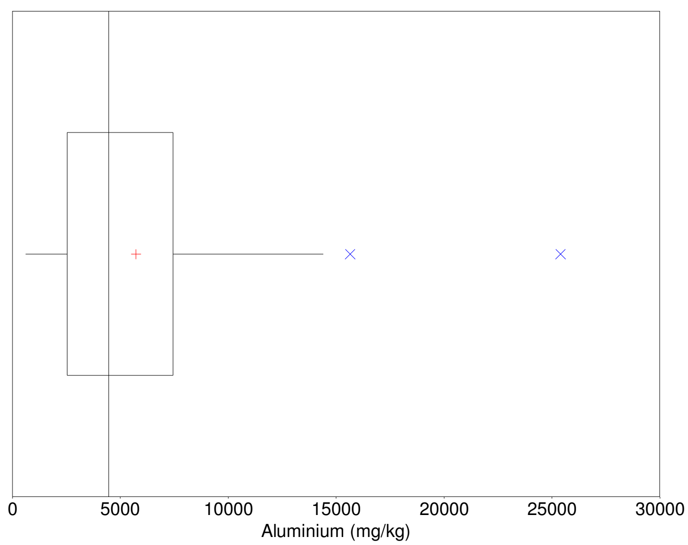
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

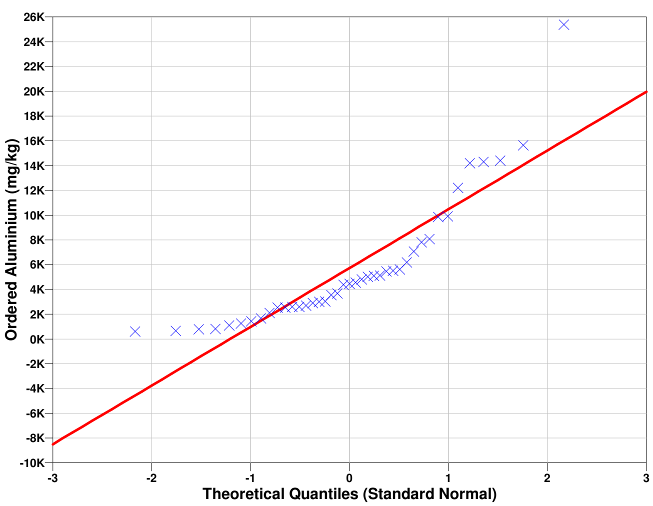
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Aluminium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.8145
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	7088
95% Non-Parametric (Chebyshev) UCL	9250

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (9250) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (6521),
SE is the standard error = (standard deviation) / (square root of n).

This *t* was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-0.98298	1.6839	Cannot Reject

The test did not reject the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean exceeds the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
30	26	Reject

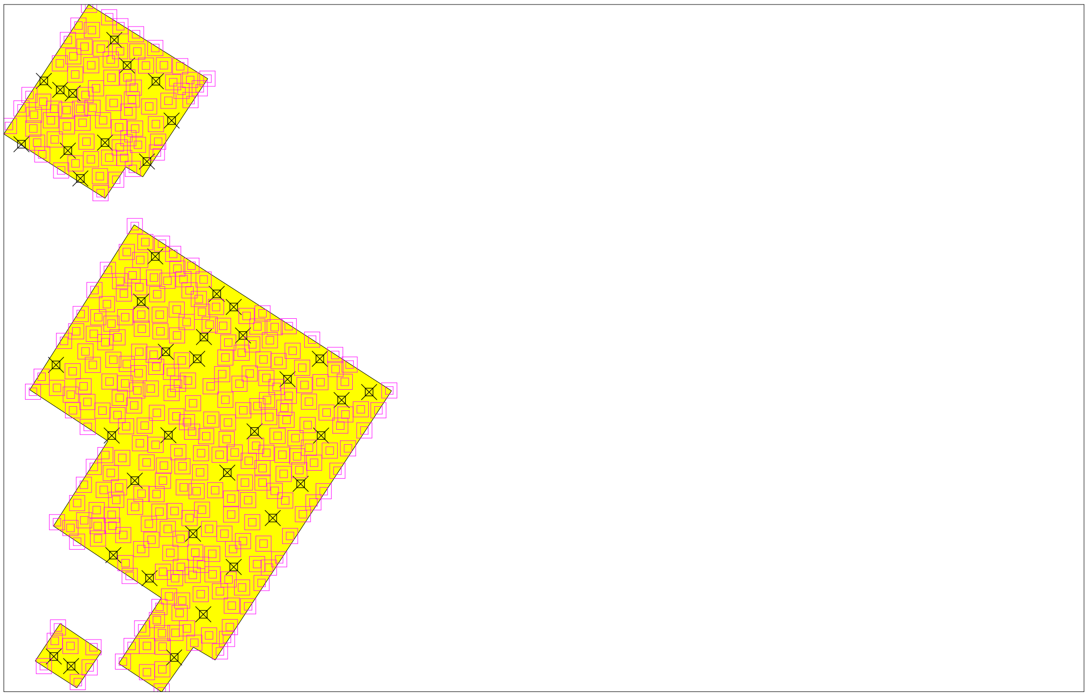
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field is also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	365
Number of samples on map [a]	365
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$183,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability (1-β) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

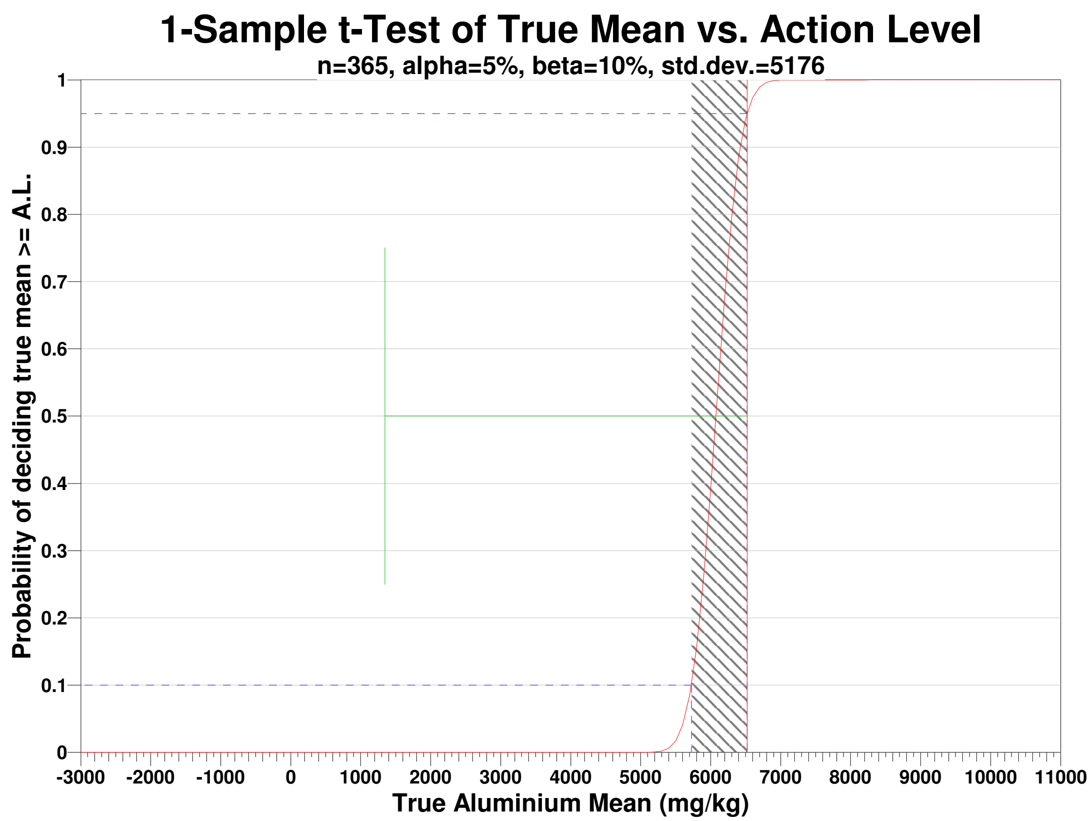
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Aluminium	365	5176 mg/kg	795 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=6521		α=5		α=10		α=15	
		s=10352	s=5176	s=10352	s=5176	s=10352	s=5176
LBGR=90	β=5	2729	684	2160	541	1813	454
	β=10	2160	541	1657	415	1355	340
	β=15	1814	455	1355	340	1084	272
LBGR=80	β=5	684	172	541	136	454	114
	β=10	541	137	415	105	340	86
	β=15	455	115	340	86	272	69
LBGR=70	β=5	305	78	241	61	202	51
	β=10	242	62	185	47	151	39
	β=15	203	52	152	39	121	31

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling

The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $183,500.00, which averages out to a per sample cost of $502.74. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**365 Samples**
Field collection costs		$100.00	$36,500.00
Analytical costs	$400.00	$400.00	$146,000.00
Sum of Field & Analytical costs		**$500.00**	**$182,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$183,500.00**

Data Analysis for Aluminium

SUMMARY STATISTICS for Aluminium								
n	365							
Min	0							
Max	25400							
Range	25400							
Mean	1176.8							
Median	0							
Variance	1.1073e+007							
StdDev	3327.6							
Std Error	174.17							
Skewness	4.0233							
Interquartile Range	0							
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0	0	0	0	0	0	4570	7830	1.565e+004

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Aluminium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.801	3.05	Yes

The test statistic 3.801 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Aluminium	
1	25400

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Lilliefors test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Lilliefors Test Statistic	0.8538
Lilliefors 5% Critical Value	0.94

The calculated Lilliefors test statistic is less than the 5% Lilliefors critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed

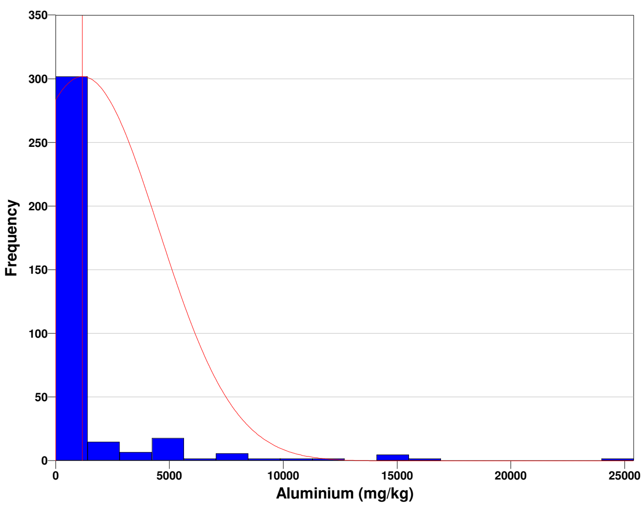
data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

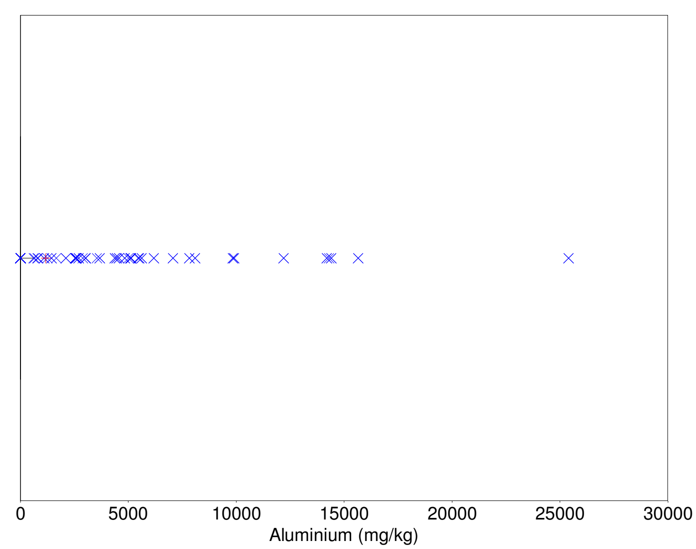
Data Plots for Aluminium
Graphical displays of the data are shown below.

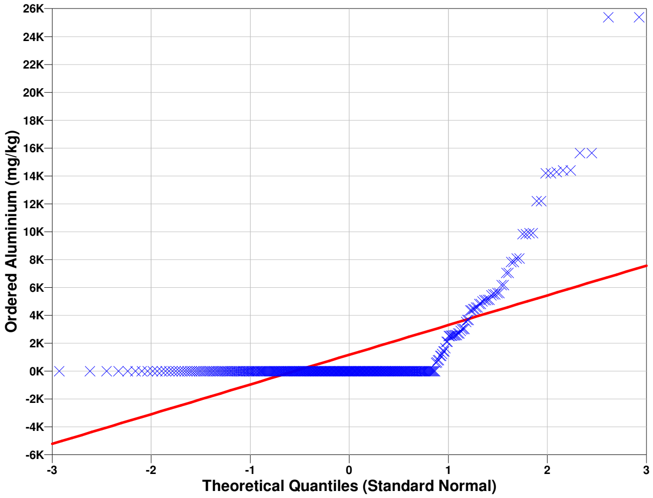
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25th and 75th percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The pth quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Aluminium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Lilliefors test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Lilliefors Test Statistic	0.4464
Lilliefors 5% Critical Value	0.04638

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	1464
95% Non-Parametric (Chebyshev) UCL	1936

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (1936) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=365 data,
AL is the action level or threshold (6521),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=364 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-30.683	1.6491	Reject

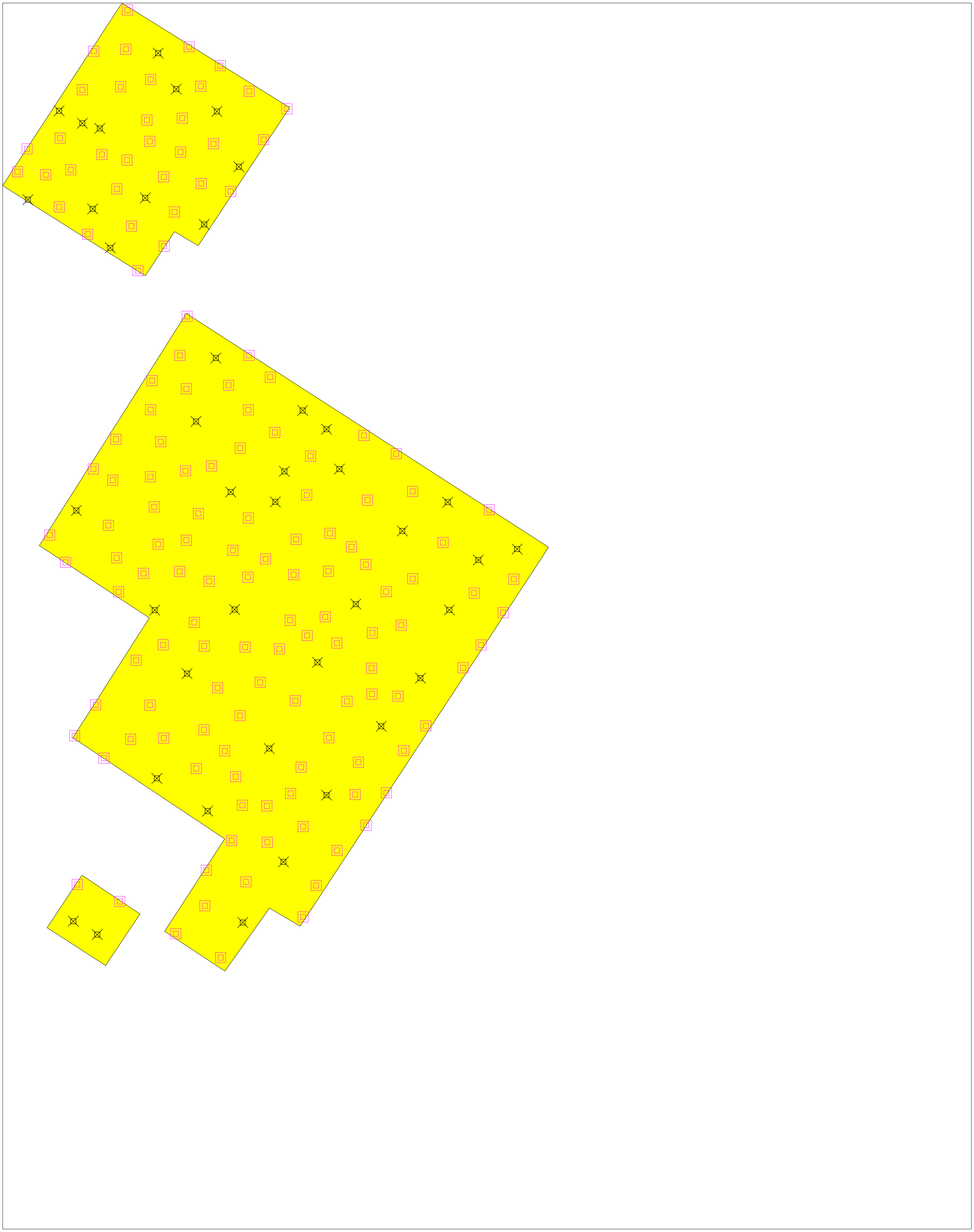
The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

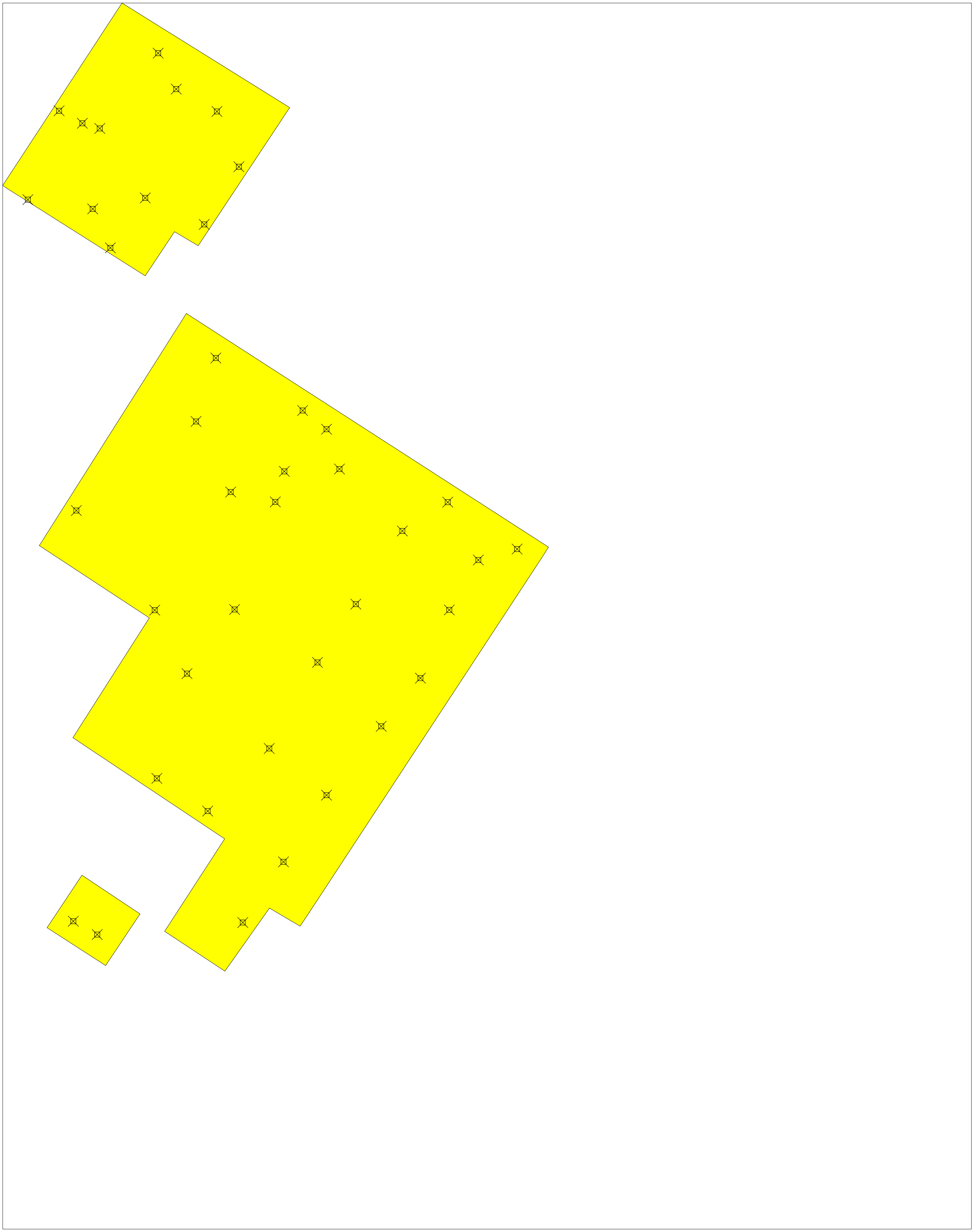
Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
344	199	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.





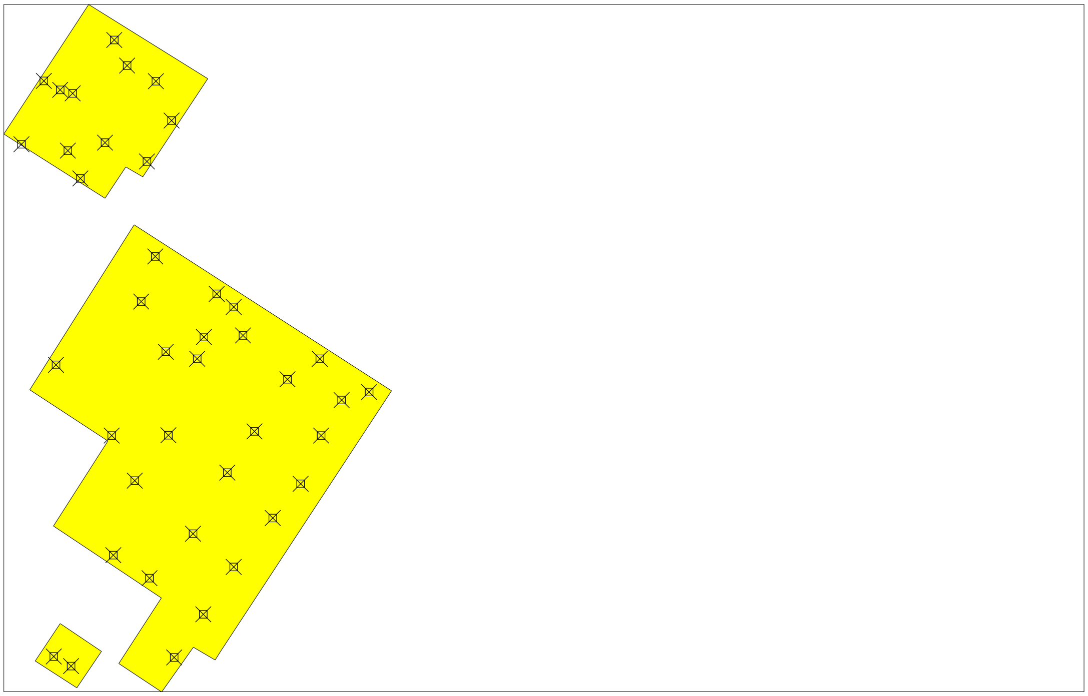
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	11
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$6,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	2.6	Manual	T
679301.1600	3083254.0340	J-12S	0.57	Manual	T
679171.2970	3083289.7960	J-04S	0.23	Manual	T
679155.0740	3083294.6960	J-03S	0.115	Manual	T
679133.4290	3083306.3130	J-01S	0.51	Manual	T
679104.2450	3083223.2620	J-02S	0.09	Manual	T
679164.8060	3083214.7100	J-06S	0.12	Manual	T
679181.2750	3083178.2880	J-08S	0.83	Manual	T
679213.7730	3083224.9730	J-09S	1.2	Manual	T
679280.5440	3083305.6810	J-10S	0.31	Manual	T
679242.7260	3083326.5280	J-07S	2	Manual	T
679225.8560	3083359.9740	J-05S	0.29	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	2.6	Manual	T
679335.0020	3082941.1720	J-21S	0.93	Manual	T
679343.5810	3082969.5980	J-19S	1.8	Manual	T
679394.8070	3082971.8300	J-24S	0.35	Manual	T
679382.8640	3083009.1130	J-20S	0.86	Manual	T
679293.5600	3082950.4980	J-17S	2	Manual	T
679360.5700	3083026.4980	J-18S	0.58	Manual	T

679297.0010	3082840.6970	J-23S	1.7	Manual	T
679252.7130	3082781.0290	J-22S	1.035	Manual	T
679222.6340	3082840.1720	J-16S	0.72	Manual	T
679279.6830	3083075.4290	J-14S	0.66	Manual	T
679149.4920	3082933.0980	J-13S	0.22	Manual	T
679272.0040	3082652.6750	J-28S	1.1	Manual	T
679224.5850	3082683.1400	J-26S	0.27805	Manual	T
679146.6460	3082549.7640	J-25S	0.305	Manual	T
679169.0760	3082537.3510	J-27S	0.53	Manual	T
679342.7410	3082605.3190	J-35S	2	Manual	T
679304.6530	3082548.6880	J-34S	3	Manual	T
679382.8900	3082667.5270	J-36S	1.45	Manual	T
679329.4380	3082711.0960	J-29S	1.3	Manual	T
679374.4420	3082791.3300	J-30S	3.1	Manual	T
679453.4760	3082914.1150	J-32S	2.5	Manual	T
679410.1490	3082845.8460	J-31S	1.6	Manual	T
679560.6070	3082897.2580	J-41S	1.3	Manual	T
679495.8840	3082940.9730	J-33S	1.4	Manual	T
679524.3310	3082886.8990	J-40S	2.8	Manual	T
679497.3310	3082840.3960	J-39S	1.5	Manual	T
679470.3570	3082776.7350	J-38S	2.4	Manual	T
679433.9450	3082731.6820	J-37S	2.2	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability (1-β) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is 1-α,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is 1-β.

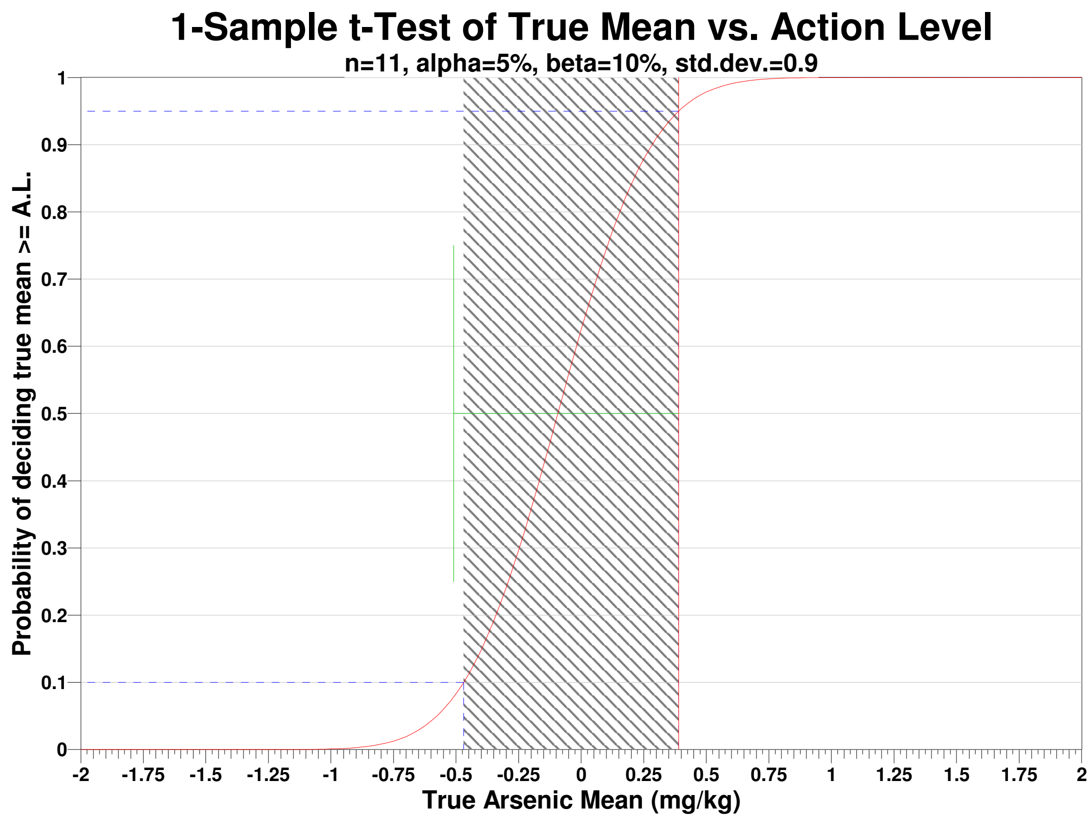
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	Δ	α	β	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Arsenic	11	0.9 mg/kg	0.86 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at 1-α on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at 1-α. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=0.39		α=5		α=10		α=15	
		s=1.8	s=0.9	s=1.8	s=0.9	s=1.8	s=0.9
LBGR=90	β=5	23055	5765	18244	4562	15316	3830
	β=10	18244	4562	13996	3500	11447	2862
	β=15	15316	3830	11447	2863	9154	2289
LBGR=80	β=5	5765	1443	4562	1141	3830	958
	β=10	4562	1142	3500	876	2862	716
	β=15	3830	959	2863	717	2289	573
LBGR=70	β=5	2563	642	2028	508	1703	426

	β=10	2029	509	1556	390	1273	319
	β=15	1703	427	1273	319	1018	255

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $6,500.00, which averages out to a per sample cost of $590.91. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**11 Samples**
Field collection costs		$100.00	$1,100.00
Analytical costs	$400.00	$400.00	$4,400.00
Sum of Field & Analytical costs		**$500.00**	**$5,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$6,500.00**

Data Analysis for Arsenic
The following data points were entered by the user for analysis.

Arsenic (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.09	0.115	0.12	0.22	0.23	0.2781	0.29	0.305	0.31	0.35
10	0.51	0.53	0.57	0.58	0.66	0.72	0.83	0.86	0.93	1.035
20	1.1	1.2	1.3	1.3	1.4	1.45	1.5	1.6	1.7	1.8
30	2	2	2	2.2	2.4	2.5	2.6	2.6	2.8	3
40	3.1									

SUMMARY STATISTICS for Arsenic	
n	41
Min	0.09
Max	3.1
Range	3.01
Mean	1.2459
Median	1.1
Variance	0.80844
StdDev	0.89913
Std Error	0.14042
Skewness	0.50263

Interquartile Range				1.57				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.09	0.1155	0.222	0.43	1.1	2	2.6	2.98	3.1

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Arsenic			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	2.062	3.05	No

None of the test statistics exceeded the corresponding critical values, therefore none of the 1 tests are significant and we conclude that at the 5% significance level there are no outliers in the data.

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.9198
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

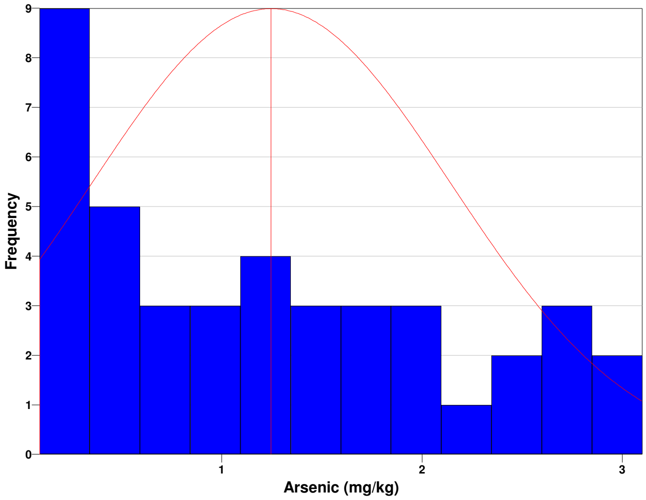
Data Plots for Arsenic
Graphical displays of the data are shown below.

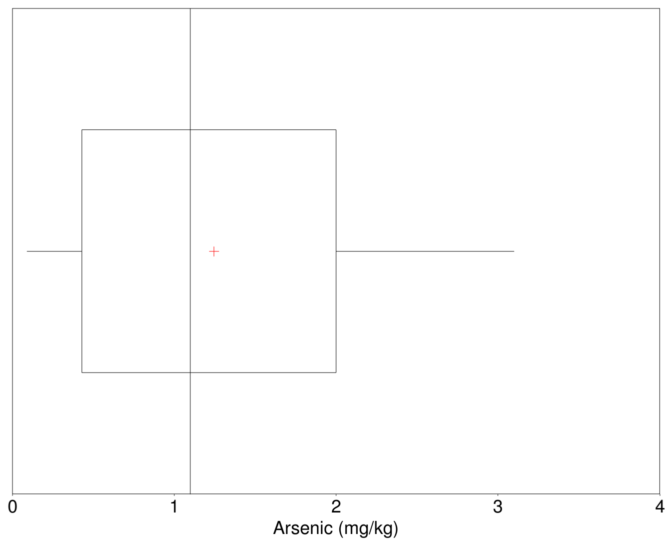
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

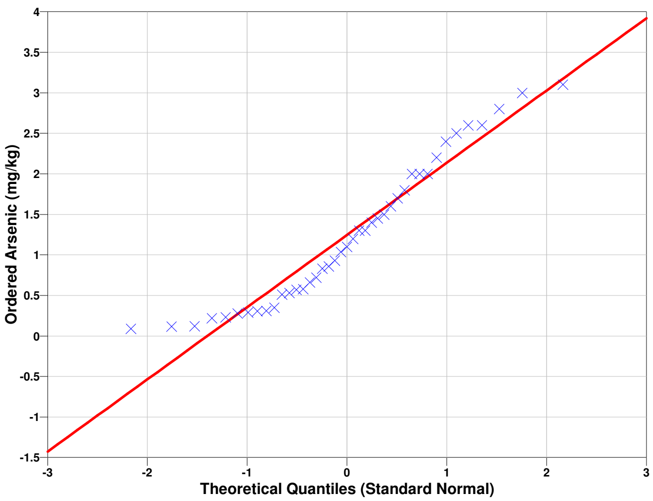
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the

distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Arsenic

A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.9178
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	1.482
95% Non-Parametric (Chebyshev) UCL	1.858

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (1.858) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (0.39),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
6.0954	1.6839	Cannot Reject

The test did not reject the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean exceeds the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
10	26	Cannot Reject
Note: There may not be enough data to reject the null hypothesis (and conclude site is clean) with 95% confidence using the MARSSIM sign test.		

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

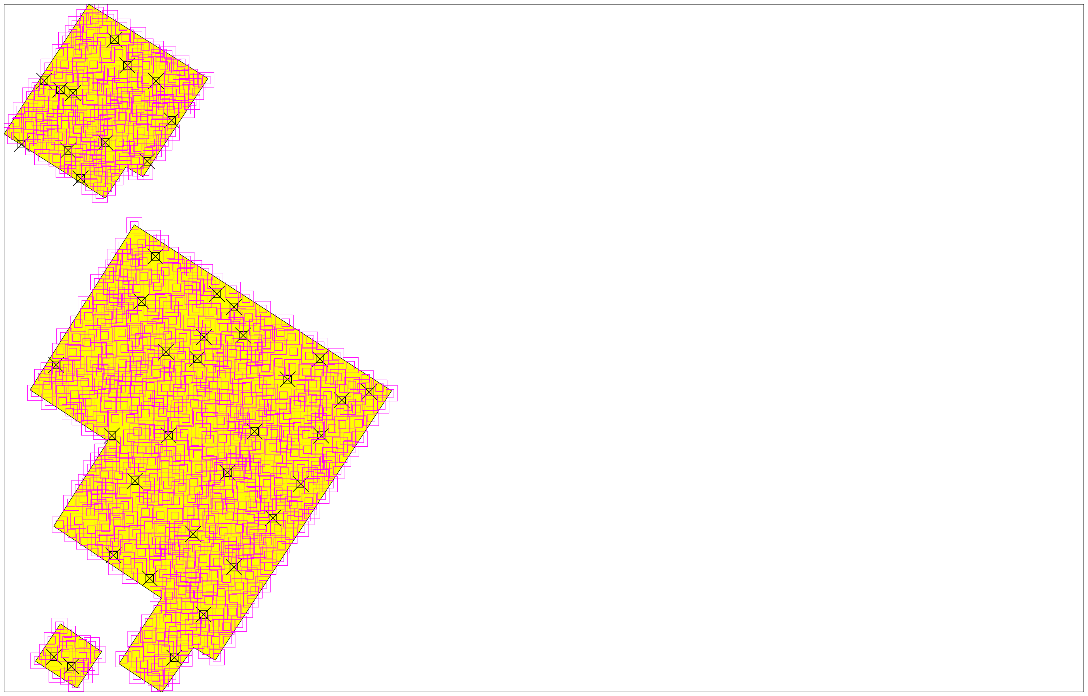
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field is also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	741
Number of samples on map [a]	741
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$371,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability (1-β) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is 1-α,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is 1-β.

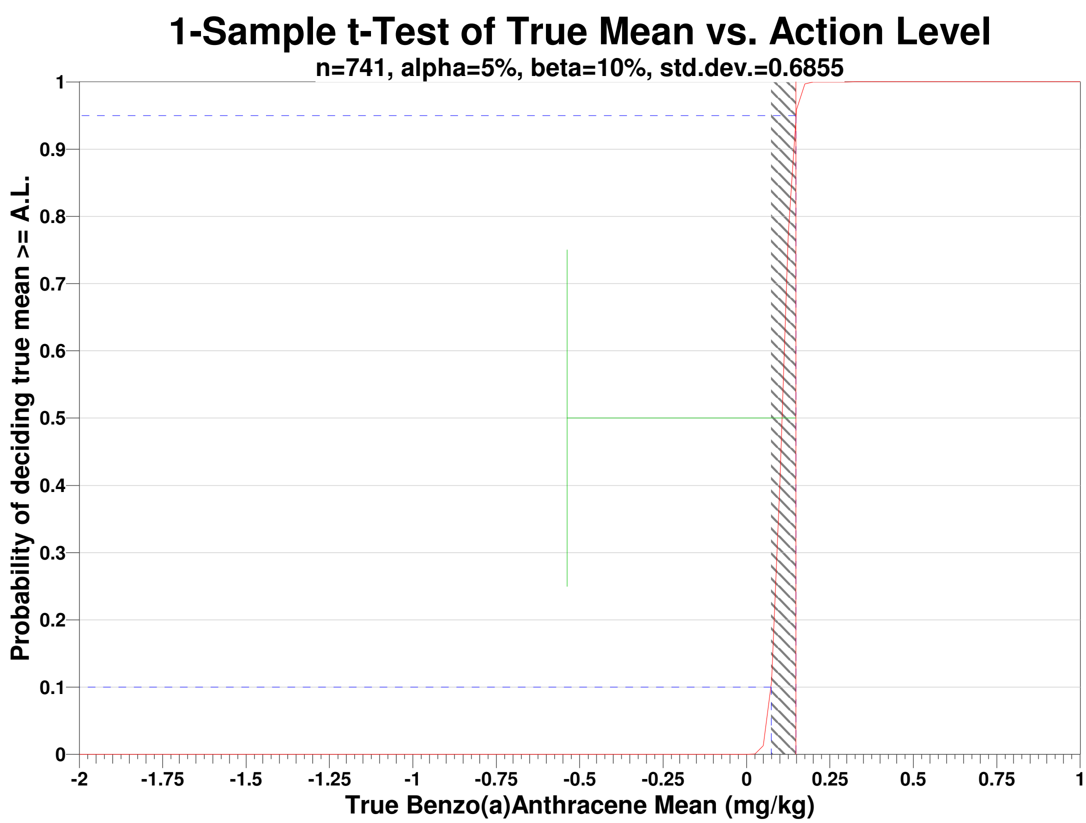
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	Δ	α	β	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Benzo(a)Anthracene	741	0.6855 mg/kg	0.0738 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at 1-α on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at 1-α. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=0.1476		**α=5**		**α=10**		**α=15**	
		s=1.371	**s=0.6855**	**s=1.371**	**s=0.6855**	**s=1.371**	**s=0.6855**
LBGR=90	**β=5**	93374	23345	73889	18473	62029	15508
	β=10	73889	18474	56682	14171	46359	11591
	β=15	62030	15509	46359	11591	37073	9269
LBGR=80	**β=5**	23345	5838	18473	4619	15508	3878
	β=10	18474	4620	14171	3544	11591	2898
	β=15	15509	3879	11591	2899	9269	2318
LBGR=70	**β=5**	10377	2596	8211	2054	6893	1724
	β=10	8212	2054	6299	1576	5152	1289
	β=15	6894	1725	5152	1289	4120	1031

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $371,500.00, which averages out to a per sample cost of $501.35. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**741 Samples**
Field collection costs		$100.00	$74,100.00
Analytical costs	$400.00	$400.00	$296,400.00
Sum of Field & Analytical costs		**$500.00**	**$370,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$371,500.00**

Data Analysis for Benzo(a)Anthracene

SUMMARY STATISTICS for Benzo(a)Anthracene								
n	741							
Min	0							
Max	3.97							
Range	3.97							
Mean	0.017023							
Median	0							
Variance	0.029269							
StdDev	0.17108							
Std Error	0.0062848							
Skewness	19.238							
Interquartile Range	0							
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0	0	0	0	0	0	0	0.039	0.3816

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Benzo(a)Anthracene			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	23.11	3.943	Yes

The test statistic 23.11 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Benzo(a)Anthracene	
1	3.97

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Lilliefors test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Lilliefors Test Statistic	0.4583
Lilliefors 5% Critical Value	0.03257

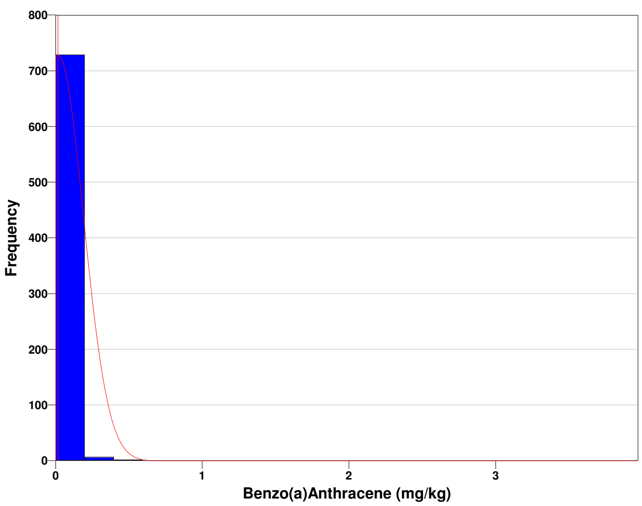
The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

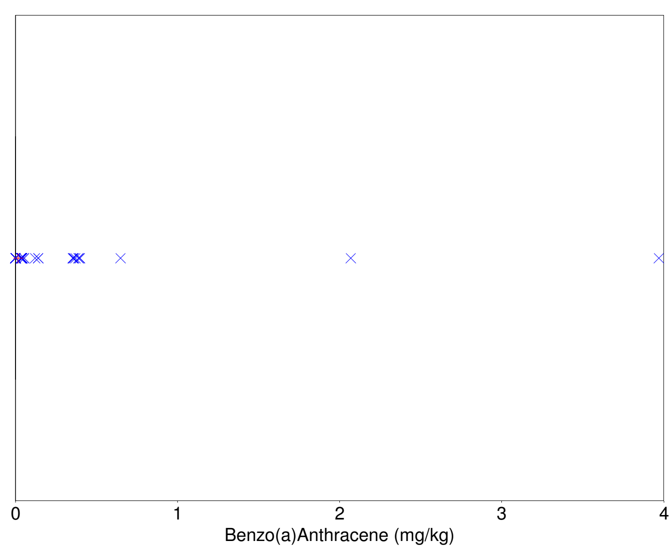
Data Plots for Benzo(a)Anthracene
Graphical displays of the data are shown below.

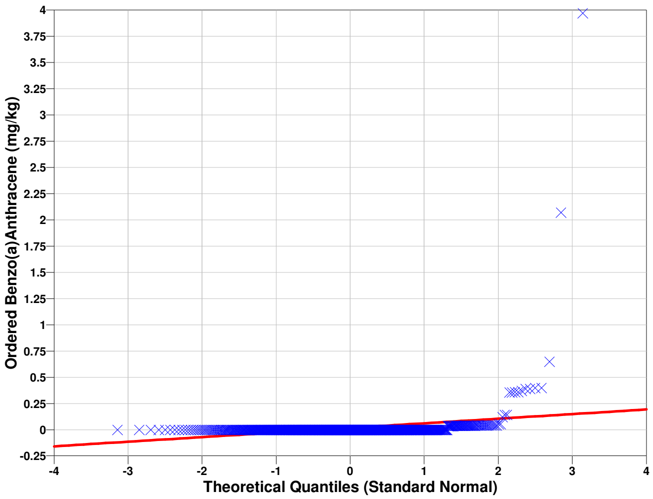
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Benzo(a)Anthracene
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Lilliefors test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Lilliefors Test Statistic	0.4604
Lilliefors 5% Critical Value	0.03255

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.02737
95% Non-Parametric (Chebyshev) UCL	0.04442

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.04442) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=741 data,
AL is the action level or threshold (0.1476),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=740 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-20.777	1.6469	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
729	393	Reject

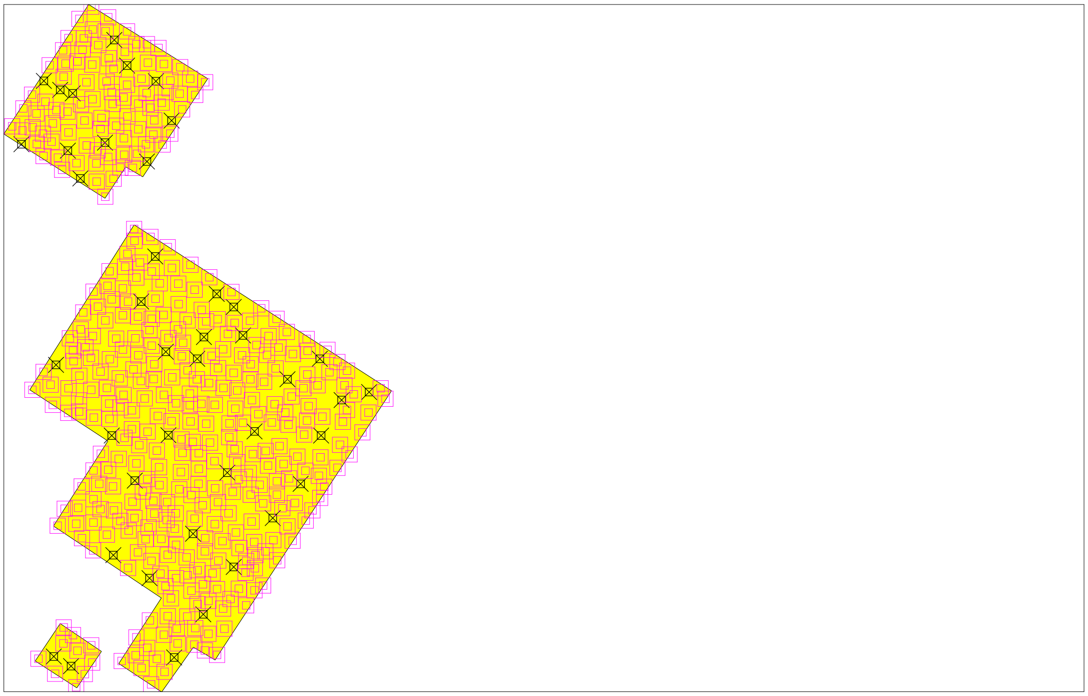
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field is also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	430
Number of samples on map [a]	430
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$216,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5 Z_{1-\alpha}^2$$

where

n is the number of samples,

S is the estimated standard deviation of the measured values including analytical error,

Δ is the width of the gray region,

α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,

β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,

$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,

$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

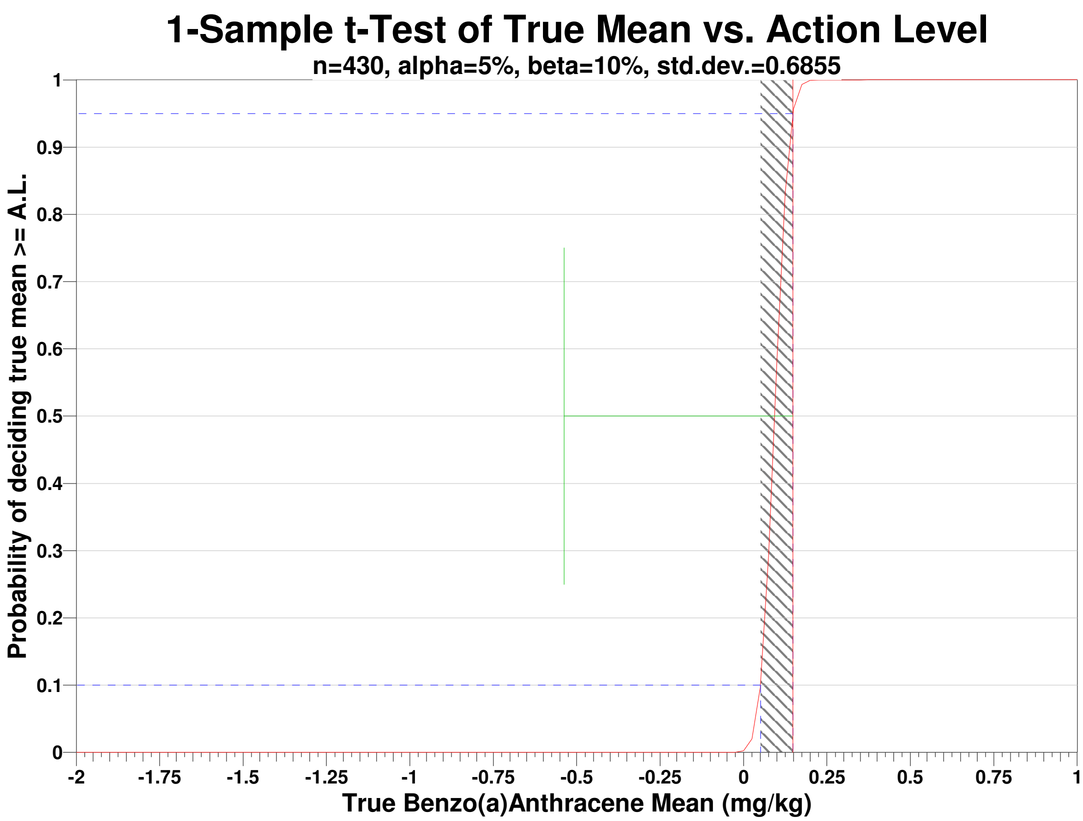
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	**n**	**Parameter**					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Benzo(a)Anthracene	430	0.6855 mg/kg	0.097 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=0.1476		α=5		α=10		α=15	
		s=1.371	s=0.6855	s=1.371	s=0.6855	s=1.371	s=0.6855
LBGR=90	β=5	93374	23345	73889	18473	62029	15508
	β=10	73889	18474	56682	14171	46359	11591
	β=15	62030	15509	46359	11591	37073	9269
LBGR=80	β=5	23345	5838	18473	4619	15508	3878
	β=10	18474	4620	14171	3544	11591	2898
	β=15	15509	3879	11591	2899	9269	2318
LBGR=70	β=5	10377	2596	8211	2054	6893	1724
	β=10	8212	2054	6299	1576	5152	1289
	β=15	6894	1725	5152	1289	4120	1031

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling

The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is \$216,000.00, which averages out to a per sample cost of \$502.33. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**430 Samples**
Field collection costs		\$100.00	\$43,000.00
Analytical costs	\$400.00	\$400.00	\$172,000.00
Sum of Field & Analytical costs		**\$500.00**	**\$215,000.00**
Fixed planning and validation costs			\$1,000.00
Total cost			**\$216,000.00**

Data Analysis for Benzo(a)Anthracene

SUMMARY STATISTICS for Benzo(a)Anthracene								
n				430				
Min				0				
Max				3.97				
Range				3.97				
Mean				0.029335				
Median				0				
Variance				0.050125				
StdDev				0.22389				
Std Error				0.010797				
Skewness				14.67				
Interquartile Range				0				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0	0	0	0	0	0	0.038	0.042	0.3975

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Benzo(a)Anthracene			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	5.434	3.05	Yes

The test statistic 5.434 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Benzo(a)Anthracene	
1	3.97

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Lilliefors test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Lilliefors Test Statistic	0.3865
Lilliefors 5% Critical Value	0.94

The calculated Lilliefors test statistic is less than the 5% Lilliefors critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed

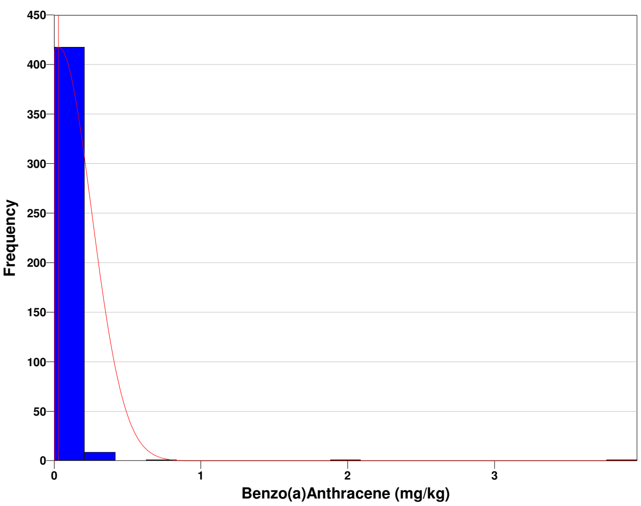
data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

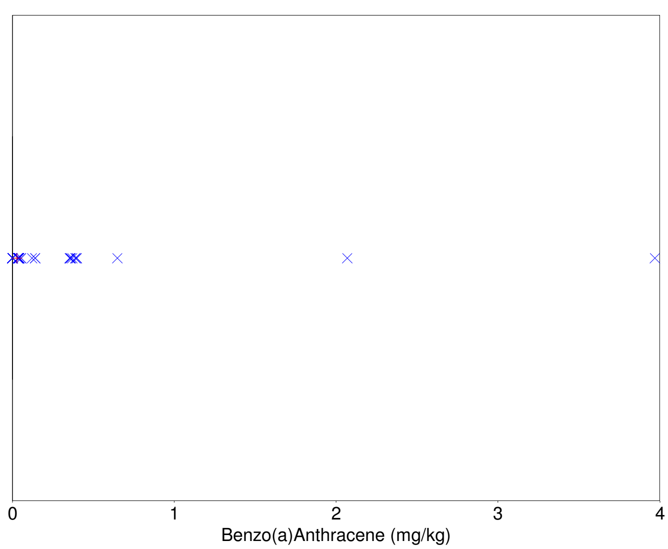
Data Plots for Benzo(a)Anthracene
Graphical displays of the data are shown below.

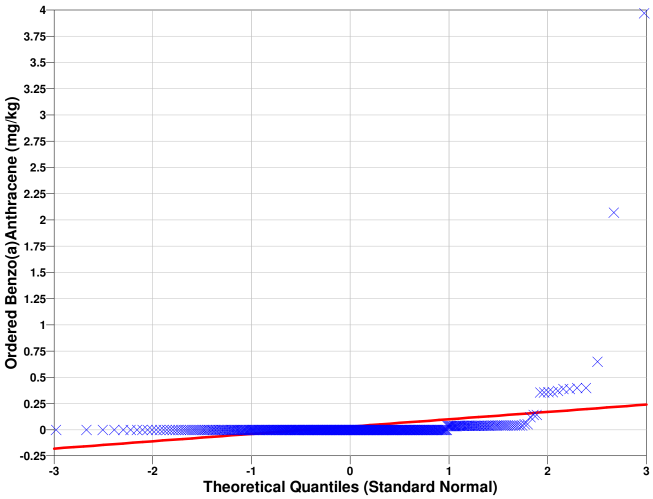
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25th and 75th percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The pth quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Benzo(a)Anthracene

A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Lilliefors test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Lilliefors Test Statistic	0.4479
Lilliefors 5% Critical Value	0.04273

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean

Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.04713
95% Non-Parametric (Chebyshev) UCL	0.0764

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.0764) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test

A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where

x is the sample mean of the n=430 data,

AL is the action level or threshold (0.1476),

SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=429 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-10.954	1.6484	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
418	233	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

Random sampling locations for comparing a mean with a fixed threshold (parametric)

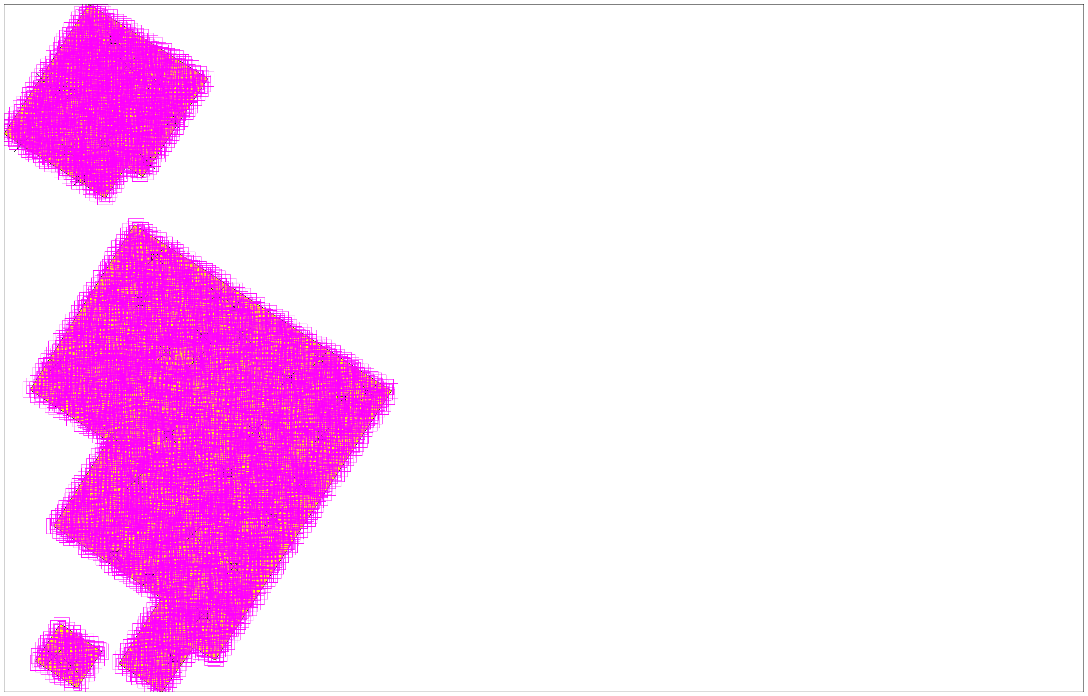
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field is also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	5055
Number of samples on map [a]	5055
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,528,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability (1-β) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

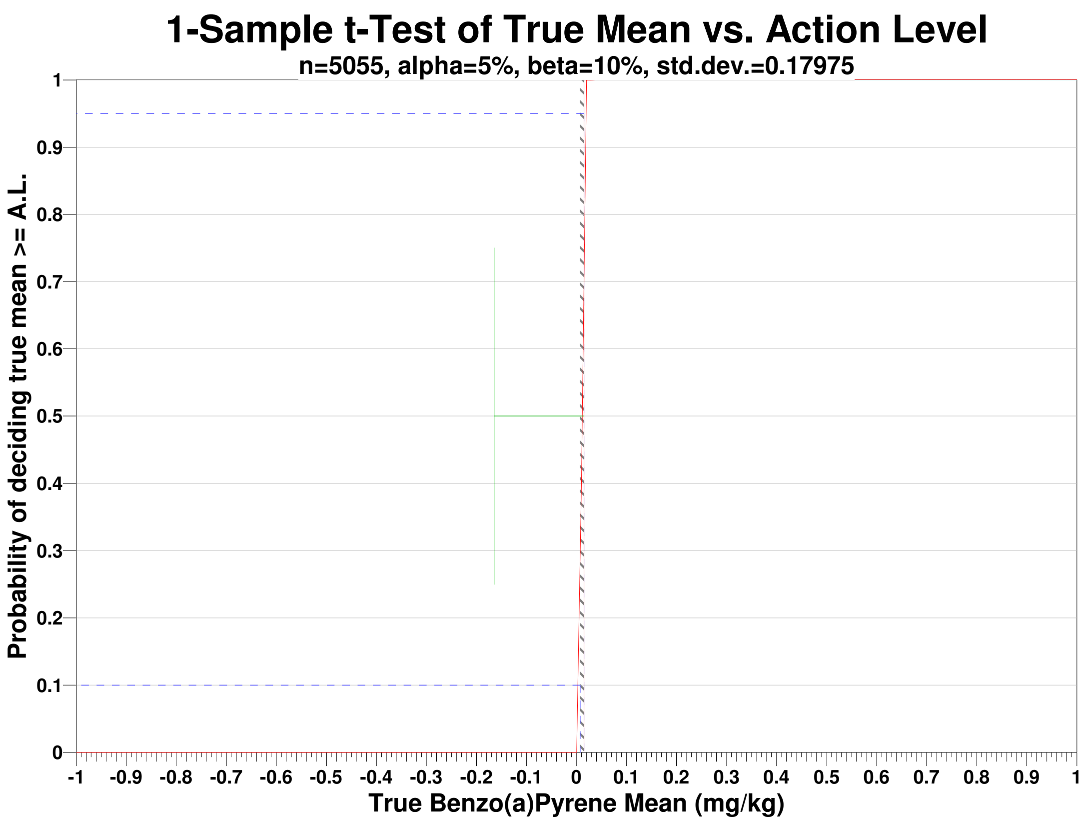
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	**n**	**Parameter**					
		S	Δ	α	β	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Benzo(a)Pyrene	5055	0.17975 mg/kg	0.0074 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=0.0148		**α=5**		**α=10**		**α=15**	
		s=0.3595	**s=0.17975**	**s=0.3595**	**s=0.17975**	**s=0.3595**	**s=0.17975**
LBGR=90	**β=5**	638543	159637	505294	126325	424192	106049
	β=10	505295	126325	387622	96906	317028	79258
	β=15	424192	106049	317028	79258	253524	63382
LBGR=80	**β=5**	159637	39911	126325	31582	106049	26513
	β=10	126325	31583	96906	24228	79258	19815
	β=15	106049	26514	79258	19815	63382	15846
LBGR=70	**β=5**	70951	17739	56145	14037	47133	11784
	β=10	56146	14038	43070	10769	35226	8807
	β=15	47134	11785	35226	8808	28170	7043

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling

The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,528,500.00, which averages out to a per sample cost of $500.20. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**5055 Samples**
Field collection costs		$100.00	$505,500.00
Analytical costs	$400.00	$400.00	$2,022,000.00
Sum of Field & Analytical costs		**$500.00**	**$2,527,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,528,500.00**

Data Analysis for Benzo(a)Pyrene

SUMMARY STATISTICS for Benzo(a)Pyrene								
n	5055							
Min	0							
Max	0.775							
Range	0.775							
Mean	0.0013544							
Median	0							
Variance	0.00044949							
StdDev	0.021201							
Std Error	0.00029819							
Skewness	25.653							
Interquartile Range	0							
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0	0	0	0	0	0	0	0	0.033

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Benzo(a)Pyrene			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.686	3.05	Yes

The test statistic 3.686 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Benzo(a)Pyrene	
1	0.775

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Lilliefors test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Lilliefors Test Statistic	0.5104
Lilliefors 5% Critical Value	0.94

The calculated Lilliefors test statistic is less than the 5% Lilliefors critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed

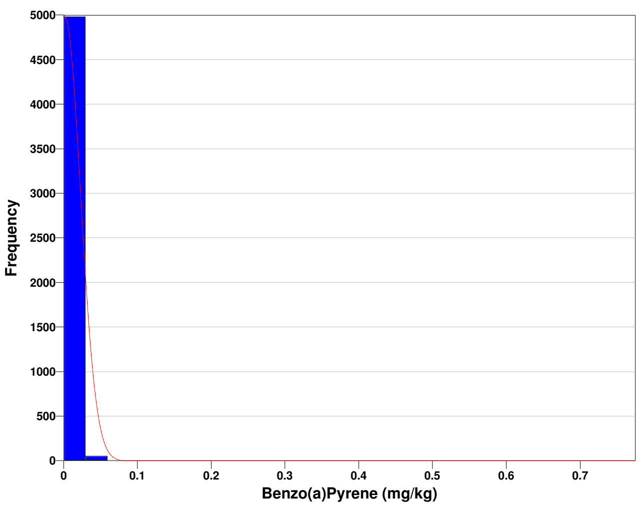
data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

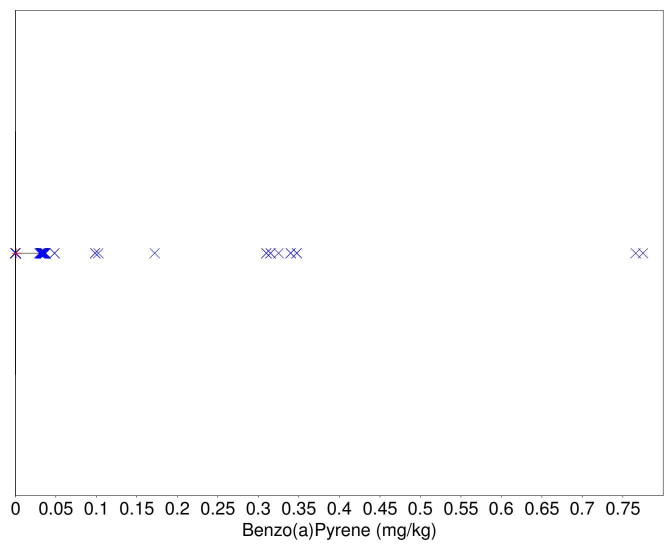
Data Plots for Benzo(a)Pyrene
Graphical displays of the data are shown below.

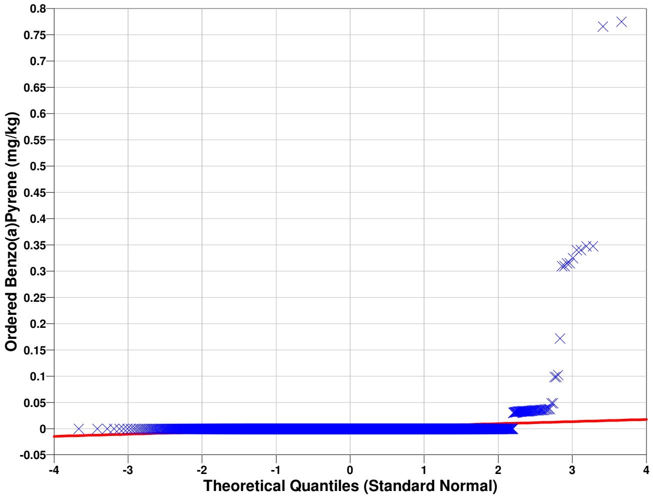
The Histogram is a plot of the fraction of the n observed data that fall within specified data "bins." A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25th and 75th percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The pth quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Benzo(a)Pyrene
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Lilliefors test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Lilliefors Test Statistic	0.5116
Lilliefors 5% Critical Value	0.01246

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.001845
95% Non-Parametric (Chebyshev) UCL	0.002654

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.002654) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=5055 data,
AL is the action level or threshold (0.0148),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=5054 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-45.09	1.6452	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
4985	2586	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

Random sampling locations for comparing a mean with a fixed threshold (parametric)

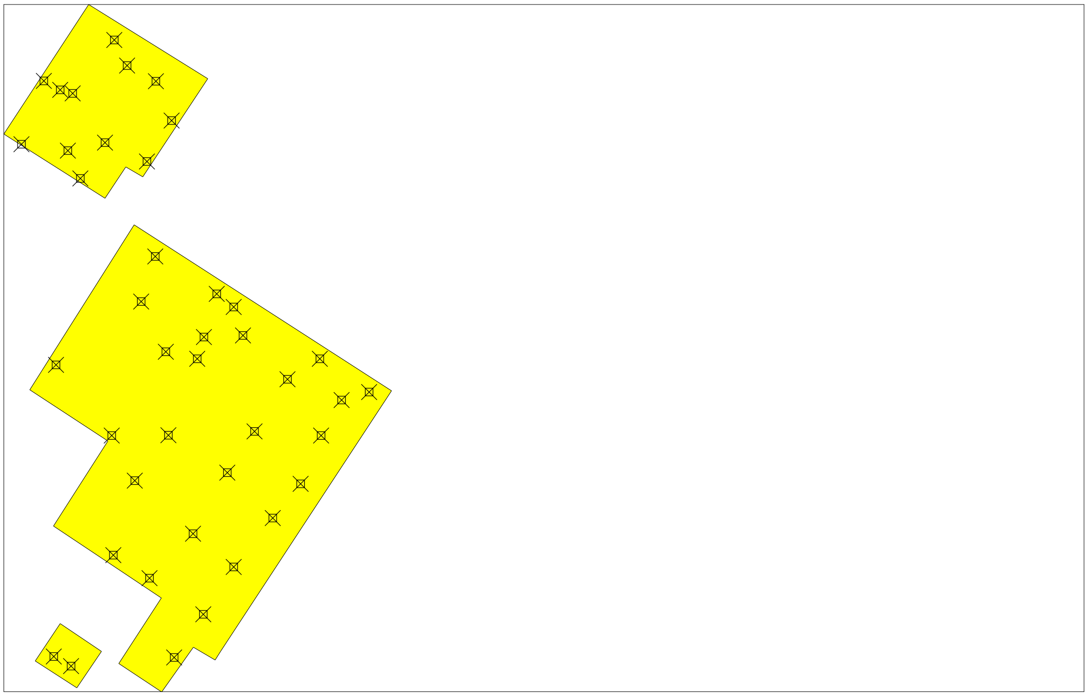
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	31
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$16,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.036	Manual	T
679301.1600	3083254.0340	J-12S	0.172	Manual	T
679171.2970	3083289.7960	J-04S	0.766	Manual	T
679155.0740	3083294.6960	J-03S	0.1025	Manual	T
679133.4290	3083306.3130	J-01S	0.037	Manual	T
679104.2450	3083223.2620	J-02S	0.031	Manual	T
679164.8060	3083214.7100	J-06S	0.03275	Manual	T
679181.2750	3083178.2880	J-08S	0.032	Manual	T
679213.7730	3083224.9730	J-09S	0.775	Manual	T
679280.5440	3083305.6810	J-10S	0.0325	Manual	T
679242.7260	3083326.5280	J-07S	0.325	Manual	T
679225.8560	3083359.9740	J-05S	0.035	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	0.033	Manual	T
679335.0020	3082941.1720	J-21S	0.34	Manual	T
679343.5810	3082969.5980	J-19S	0.3475	Manual	T
679394.8070	3082971.8300	J-24S	0.315	Manual	T
679382.8640	3083009.1130	J-20S	0.31	Manual	T
679293.5600	3082950.4980	J-17S	0.0355	Manual	T
679360.5700	3083026.4980	J-18S	0.034	Manual	T

679297.0010	3082840.6970	J-23S	0.0315	Manual	T
679252.7130	3082781.0290	J-22S	0.03275	Manual	T
679222.6340	3082840.1720	J-16S	0.0335	Manual	T
679279.6830	3083075.4290	J-14S	0.0985	Manual	T
679149.4920	3082933.0980	J-13S	0.0325	Manual	T
679272.0040	3082652.6750	J-28S	0.035	Manual	T
679224.5850	3082683.1400	J-26S	0.03325	Manual	T
679146.6460	3082549.7640	J-25S	0.035	Manual	T
679169.0760	3082537.3510	J-27S	0.0345	Manual	T
679342.7410	3082605.3190	J-35S	0.0335	Manual	T
679304.6530	3082548.6880	J-34S	0.0485	Manual	T
679382.8900	3082667.5270	J-36S	0.035	Manual	T
679329.4380	3082711.0960	J-29S	0.033	Manual	T
679374.4420	3082791.3300	J-30S	0.0375	Manual	T
679453.4760	3082914.1150	J-32S	0.037	Manual	T
679410.1490	3082845.8460	J-31S	0.03375	Manual	T
679560.6070	3082897.2580	J-41S	0.03	Manual	T
679495.8840	3082940.9730	J-33S	0.0325	Manual	T
679524.3310	3082886.8990	J-40S	0.0335	Manual	T
679497.3310	3082840.3960	J-39S	0.036	Manual	T
679470.3570	3082776.7350	J-38S	0.0315	Manual	T
679433.9450	3082731.6820	J-37S	0.0315	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

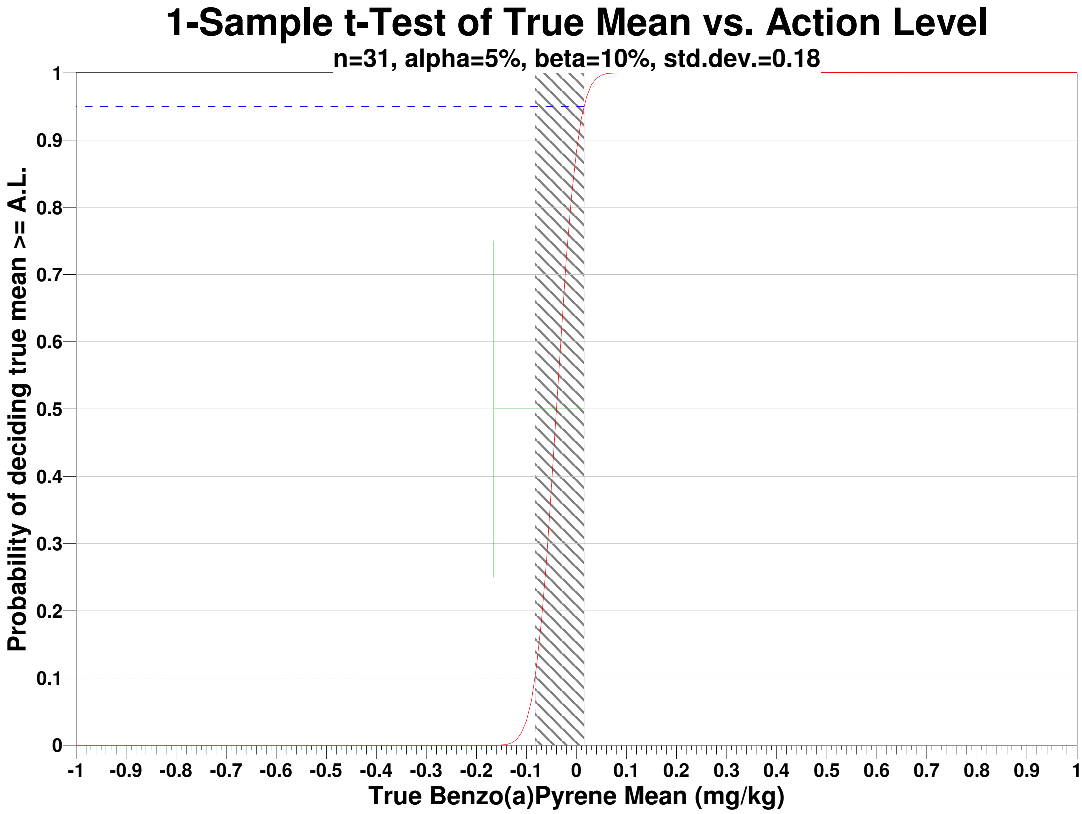
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	**n**	**Parameter**					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Benzo(a)Pyrene	31	0.18 mg/kg	0.0976756 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=0.0148		α=5		α=10		α=15	
		s=0.36	**s=0.18**	**s=0.36**	**s=0.18**	**s=0.36**	**s=0.18**
LBGR=90	β=5	640320	160081	506701	126676	425372	106344
	β=10	506701	126677	388701	97176	317910	79478
	β=15	425373	106345	317910	79479	254230	63558
LBGR=80	β=5	160081	40022	126676	31670	106344	26587
	β=10	126677	31671	97176	24295	79478	19870
	β=15	106345	26588	79479	19871	63558	15890
LBGR=70	β=5	71148	17788	56301	14076	47265	11817

	β=10	56302	14077	43190	10799	35324	8832
	β=15	47265	11818	35325	8832	28249	7063

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $16,500.00, which averages out to a per sample cost of $532.26. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**31 Samples**
Field collection costs		$100.00	$3,100.00
Analytical costs	$400.00	$400.00	$12,400.00
Sum of Field & Analytical costs		**$500.00**	**$15,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$16,500.00**

Data Analysis for Benzo(a)Pyrene
The following data points were entered by the user for analysis.

Benzo(a)Pyrene (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.03	0.031	0.0315	0.0315	0.0315	0.032	0.0325	0.0325	0.0325	0.03275
10	0.03275	0.033	0.033	0.03325	0.0335	0.0335	0.0335	0.03375	0.034	0.0345
20	0.035	0.035	0.035	0.035	0.0355	0.036	0.036	0.037	0.037	0.0375
30	0.0485	0.0985	0.1025	0.172	0.31	0.315	0.325	0.34	0.3475	0.766
40	0.775									

SUMMARY STATISTICS for Benzo(a)Pyrene	
n	41
Min	0.03
Max	0.775
Range	0.745
Mean	0.11248
Median	0.035
Variance	0.032309
StdDev	0.17975
Std Error	0.028072
Skewness	2.7405

Interquartile Range				0.04075				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.03	0.03105	0.0315	0.03275	0.035	0.0735	0.337	0.7241	0.775

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Benzo(a)Pyrene			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.686	3.05	Yes

The test statistic 3.686 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Benzo(a)Pyrene	
1	0.775

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.5104
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

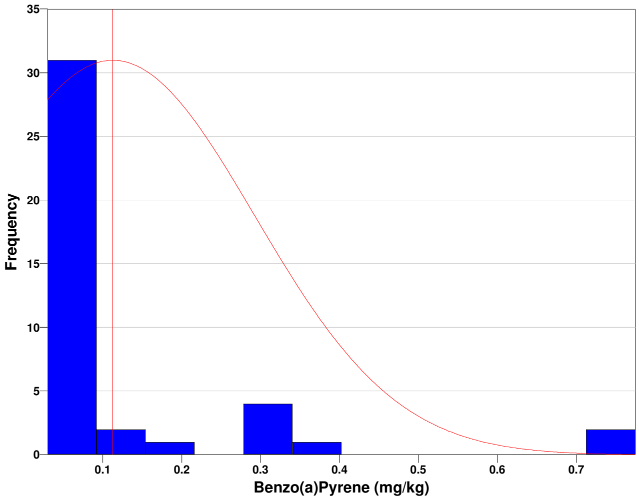
Data Plots for Benzo(a)Pyrene
Graphical displays of the data are shown below.

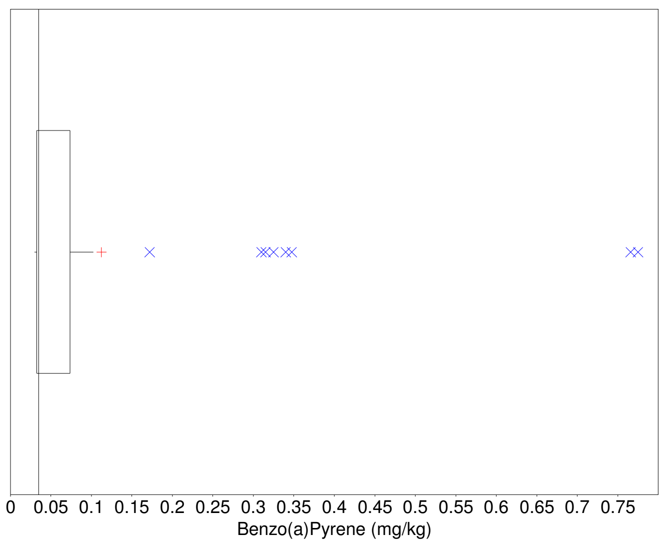
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

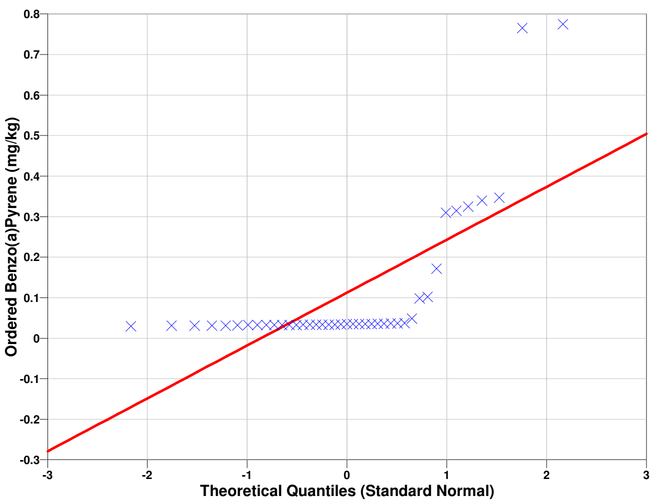
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Benzo(a)Pyrene
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.5155
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.1597
95% Non-Parametric (Chebyshev) UCL	0.2348

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.2348) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (0.0148),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
3.4795	1.6839	Cannot Reject

The test did not reject the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean exceeds the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
0	26	Cannot Reject
Note: There may not be enough data to reject the null hypothesis (and conclude site is clean) with 95% confidence using the MARSSIM sign test.		

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

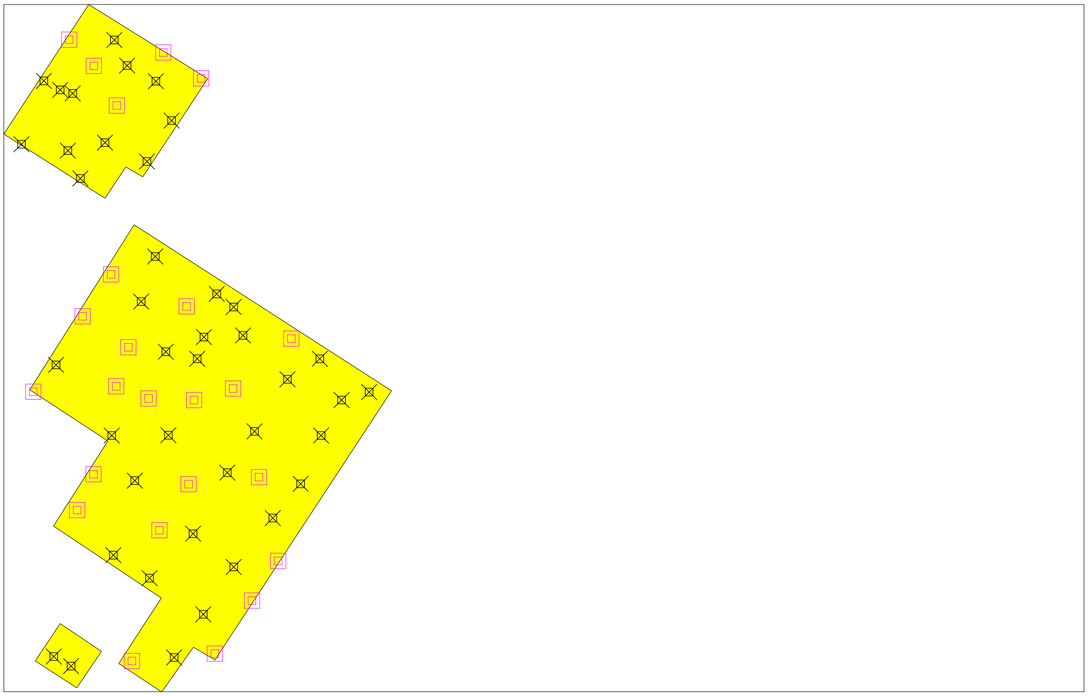
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	65
Number of samples on map [a]	65
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$33,500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.0465	Manual	T
679301.1600	3083254.0340	J-12S	0.218	Manual	T
679171.2970	3083289.7960	J-04S	0.465	Manual	T
679155.0740	3083294.6960	J-03S	0.135	Manual	T
679133.4290	3083306.3130	J-01S	0.048	Manual	T
679104.2450	3083223.2620	J-02S	0.04	Manual	T
679164.8060	3083214.7100	J-06S	0.04225	Manual	T
679181.2750	3083178.2880	J-08S	0.0415	Manual	T
679213.7730	3083224.9730	J-09S	1.03	Manual	T
679280.5440	3083305.6810	J-10S	0.042	Manual	T
679242.7260	3083326.5280	J-07S	0.42	Manual	T
679225.8560	3083359.9740	J-05S	0.045	Manual	T
679339.9414	3083309.8490	J-15S	0.043	Adaptive-Fill	
679229.3268	3083274.1226	J-21S	0.44	Adaptive-Fill	
679166.5144	3083360.6002	J-19S	0.45	Adaptive-Fill	
679199.0125	3083325.9512	J-24S	0.41	Adaptive-Fill	
679290.1895	3083343.7836	J-20S	0.405	Adaptive-Fill	

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	0.043	Manual	T
679335.0020	3082941.1720	J-21S	0.44	Manual	T

679343.5810	3082969.5980	J-19S	0.45	Manual	T
679394.8070	3082971.8300	J-24S	0.41	Manual	T
679382.8640	3083009.1130	J-20S	0.405	Manual	T
679293.5600	3082950.4980	J-17S	0.046	Manual	T
679360.5700	3083026.4980	J-18S	0.044	Manual	T
679297.0010	3082840.6970	J-23S	0.041	Manual	T
679252.7130	3082781.0290	J-22S	0.0425	Manual	T
679222.6340	3082840.1720	J-16S	0.0435	Manual	T
679279.6830	3083075.4290	J-14S	0.181	Manual	T
679149.4920	3082933.0980	J-13S	0.042	Manual	T
679272.0040	3082652.6750	J-28S	0.045	Manual	T
679224.5850	3082683.1400	J-26S	0.04275	Manual	T
679146.6460	3082549.7640	J-25S	0.04525	Manual	T
679169.0760	3082537.3510	J-27S	0.0445	Manual	T
679342.7410	3082605.3190	J-35S	0.0435	Manual	T
679304.6530	3082548.6880	J-34S	0.065	Manual	T
679382.8900	3082667.5270	J-36S	0.0455	Manual	T
679329.4380	3082711.0960	J-29S	0.0425	Manual	T
679374.4420	3082791.3300	J-30S	0.0485	Manual	T
679453.4760	3082914.1150	J-32S	0.048	Manual	T
679410.1490	3082845.8460	J-31S	0.04375	Manual	T
679560.6070	3082897.2580	J-41S	0.0385	Manual	T
679495.8840	3082940.9730	J-33S	0.042	Manual	T
679524.3310	3082886.8990	J-40S	0.043	Manual	T
679497.3310	3082840.3960	J-39S	0.0465	Manual	T
679470.3570	3082776.7350	J-38S	0.041	Manual	T
679433.9450	3082731.6820	J-37S	0.041	Manual	T
679184.4446	3082996.9264		0	Adaptive-Fill	
679177.1426	3082742.3877		0	Adaptive-Fill	
679228.5215	3082904.9556		0	Adaptive-Fill	
679381.8163	3082901.9806		0	Adaptive-Fill	
679441.0893	3082675.5360		0	Adaptive-Fill	
679358.0925	3082553.5465		0	Adaptive-Fill	
679330.4159	3082886.8562		0	Adaptive-Fill	
679221.7466	3083051.9971		0	Adaptive-Fill	
679198.5354	3082789.3512		0	Adaptive-Fill	
679249.1891	3082543.7262		0	Adaptive-Fill	
679244.2914	3082956.2636		0	Adaptive-Fill	
679406.6415	3082623.4861		0	Adaptive-Fill	
679119.5052	3082897.8954		0	Adaptive-Fill	
679271.0195	3082888.8959		0	Adaptive-Fill	

679458.5374	3082967.5965		0	Adaptive-Fill	
679285.1118	3082715.7475		0	Adaptive-Fill	
679320.7859	3083010.0230		0	Adaptive-Fill	
679323.6573	3082776.1377		0	Adaptive-Fill	
679415.7477	3082785.3452		0	Adaptive-Fill	

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

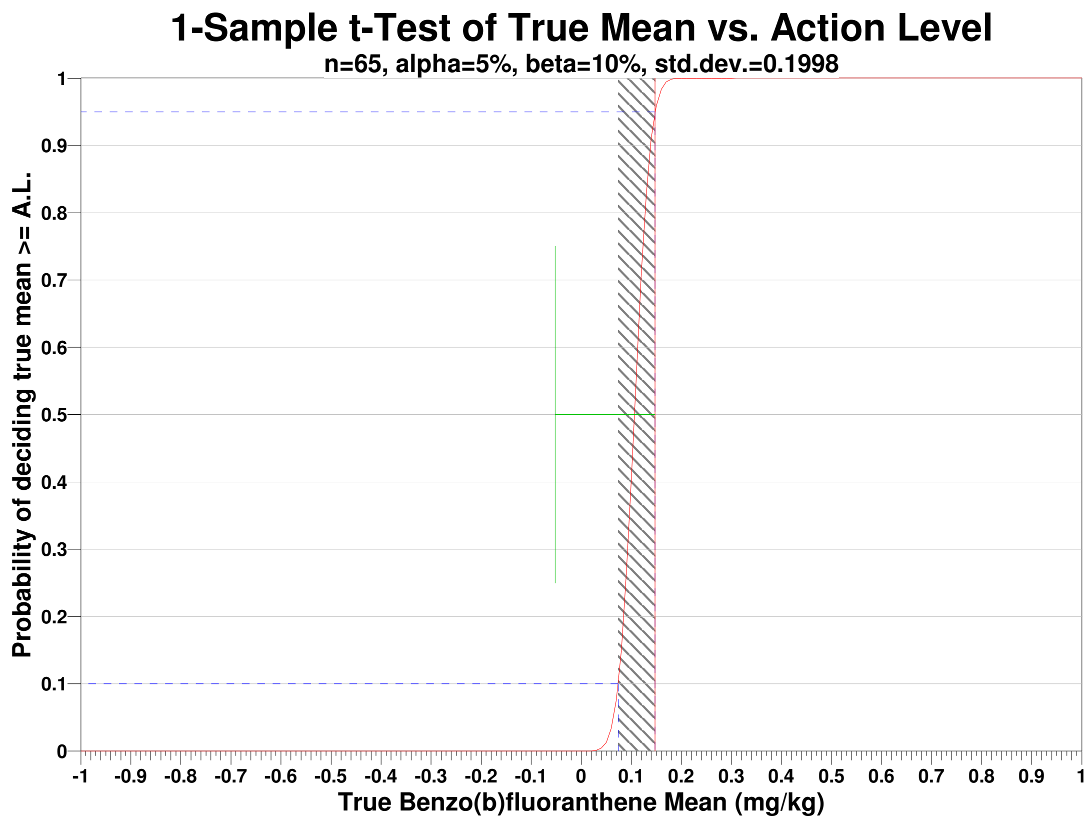
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	Δ	α	β	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Benzo(b)fluoranthene	65	0.1998 mg/kg	0.0738 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at 1-α on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at 1-α. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples

AL=0.1476		α=5		α=10		α=15	
		s=0.3996	s=0.1998	s=0.3996	s=0.1998	s=0.3996	s=0.1998
LBGR=90	β=5	7934	1985	6278	1571	5270	1318
	β=10	6279	1571	4816	1205	3939	986
	β=15	5271	1319	3940	986	3150	788
LBGR=80	β=5	1985	498	1571	394	1318	330
	β=10	1571	394	1205	302	986	247
	β=15	1319	331	986	247	788	198
LBGR=70	β=5	883	222	699	176	587	147
	β=10	699	176	536	135	439	110
	β=15	587	148	439	111	351	89

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $33,500.00, which averages out to a per sample cost of $515.38. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**65 Samples**
Field collection costs		$100.00	$6,500.00
Analytical costs	$400.00	$400.00	$26,000.00
Sum of Field & Analytical costs		**$500.00**	**$32,500.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$33,500.00**

Data Analysis for Benzo(b)fluoranthene

SUMMARY STATISTICS for Benzo(b)fluoranthene	
n	2094
Min	0
Max	1.03
Range	1.03
Mean	0.0040505
Median	0
Variance	0.001489
StdDev	0.038588
Std Error	0.00084325
Skewness	15.668

Interquartile Range				0				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0	0	0	0	0	0	0	0	0.048

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Benzo(b)fluoranthene			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	26.59	4.207	Yes

The test statistic 26.59 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Benzo(b)fluoranthene	
1	1.03

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Lilliefors test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Lilliefors Test Statistic	0.5122
Lilliefors 5% Critical Value	0.01937

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

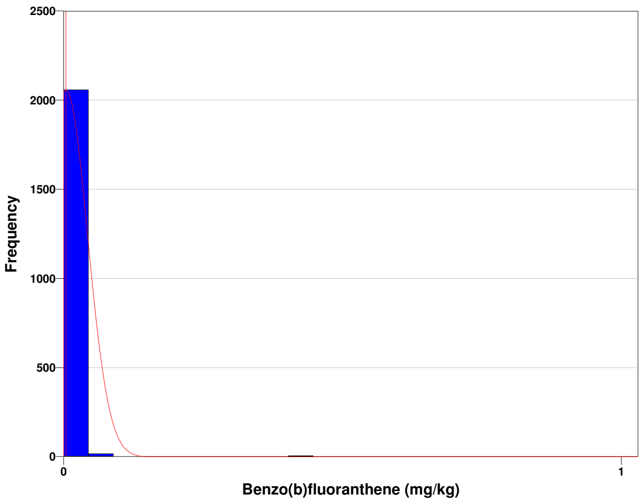
Data Plots for Benzo(b)fluoranthene
Graphical displays of the data are shown below.

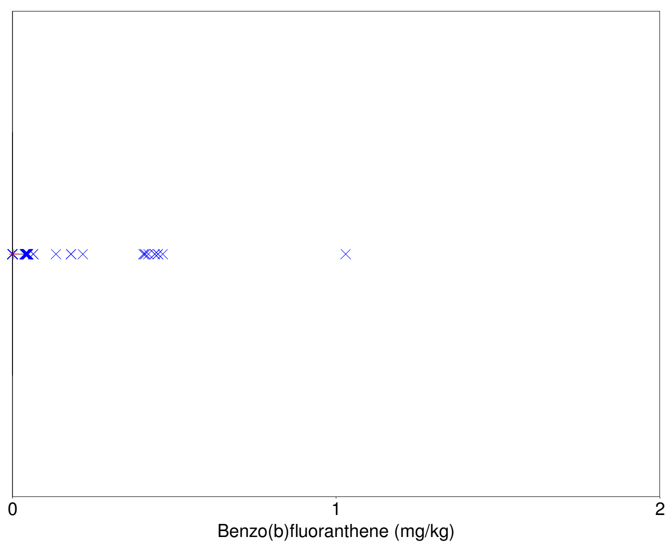
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

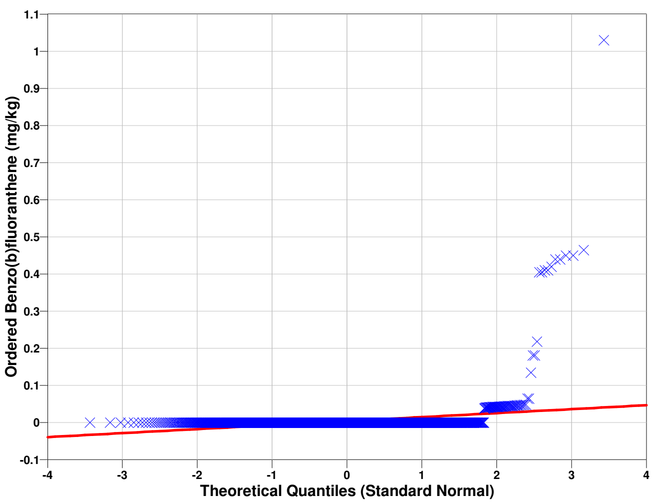
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Benzo(b)fluoranthene
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Lilliefors test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Lilliefors Test Statistic	0.5084
Lilliefors 5% Critical Value	0.01936

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.005438
95% Non-Parametric (Chebyshev) UCL	0.007726

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.007726) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=2094 data,
AL is the action level or threshold (0.1476),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=2093 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-170.23	1.6456	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
2080	1085	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

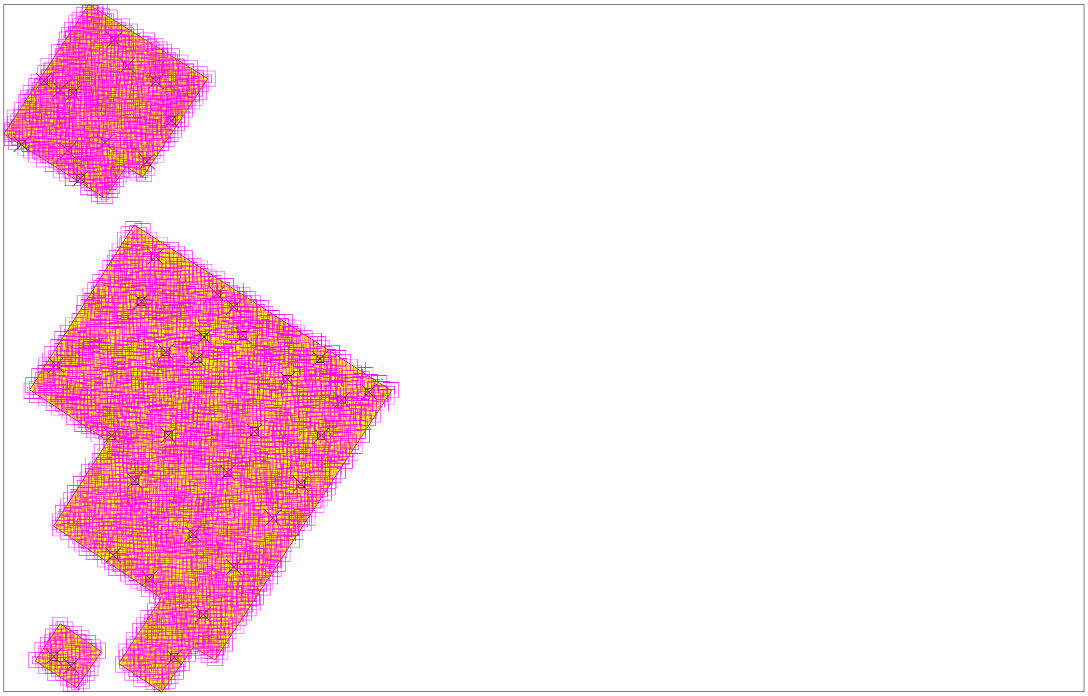
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field is also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2094
Number of samples on map [a]	2094
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$1,048,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

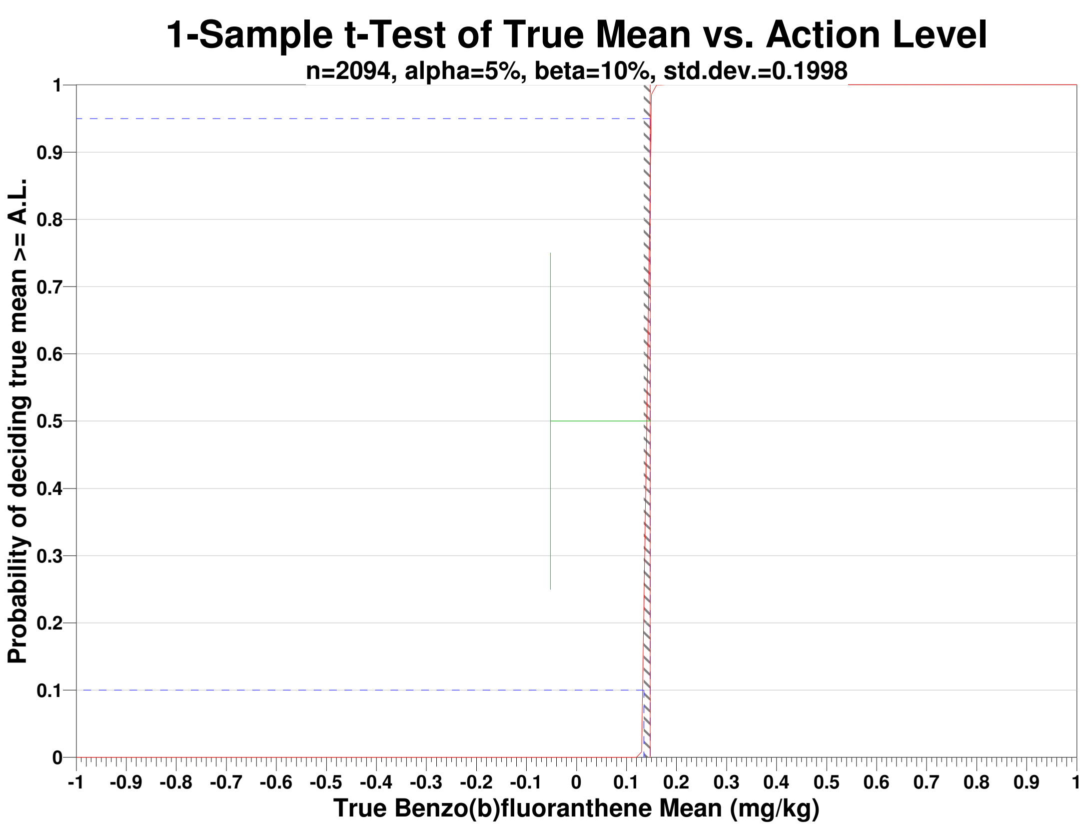
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	**n**	**Parameter**					
		S	**Δ**	**α**	**β**	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Benzo(b)fluoranthene	2094	0.1998 mg/kg	0.0127829 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=0.1476		**α=5**		**α=10**		**α=15**	
		s=0.3996	**s=0.1998**	**s=0.3996**	**s=0.1998**	**s=0.3996**	**s=0.1998**
LBGR=90	**β=5**	7934	1985	6278	1571	5270	1318
	β=10	6279	1571	4816	1205	3939	986
	β=15	5271	1319	3940	986	3150	788
LBGR=80	**β=5**	1985	498	1571	394	1318	330
	β=10	1571	394	1205	302	986	247
	β=15	1319	331	986	247	788	198
LBGR=70	**β=5**	883	222	699	176	587	147
	β=10	699	176	536	135	439	110
	β=15	587	148	439	111	351	89

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling

The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $1,048,000.00, which averages out to a per sample cost of $500.48. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2094 Samples**
Field collection costs		$100.00	$209,400.00
Analytical costs	$400.00	$400.00	$837,600.00
Sum of Field & Analytical costs		**$500.00**	**$1,047,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$1,048,000.00**

Data Analysis for Benzo(b)fluoranthene

SUMMARY STATISTICS for Benzo(b)fluoranthene								
n	2094							
Min	0							
Max	1.03							
Range	1.03							
Mean	0.0040505							
Median	0							
Variance	0.001489							
StdDev	0.038588							
Std Error	0.00084325							
Skewness	15.668							
Interquartile Range	0							
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0	0	0	0	0	0	0	0	0.048

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Benzo(b)fluoranthene			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	26.59	4.207	Yes

The test statistic 26.59 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Benzo(b)fluoranthene	
1	1.03

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Lilliefors test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Lilliefors Test Statistic	0.5122
Lilliefors 5% Critical Value	0.01937

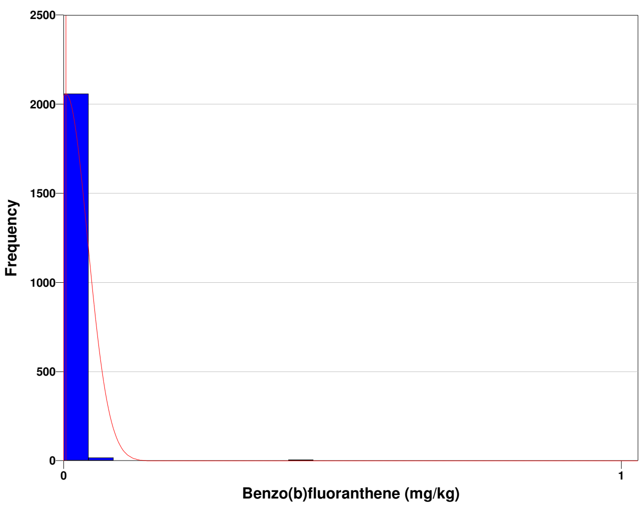
The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

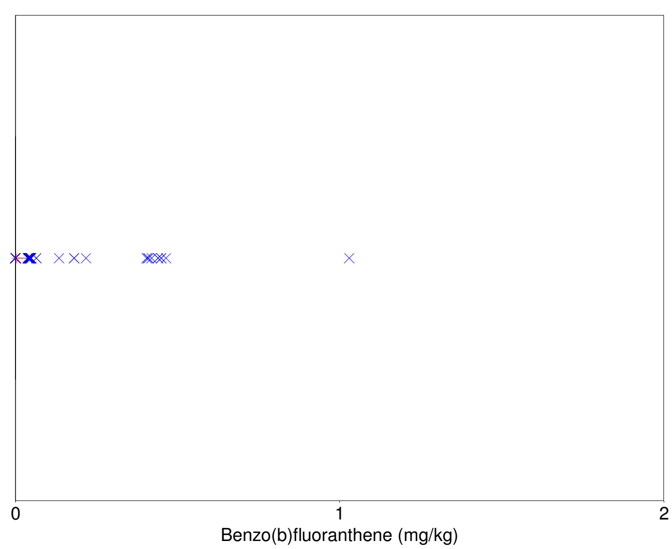
Data Plots for Benzo(b)fluoranthene
Graphical displays of the data are shown below.

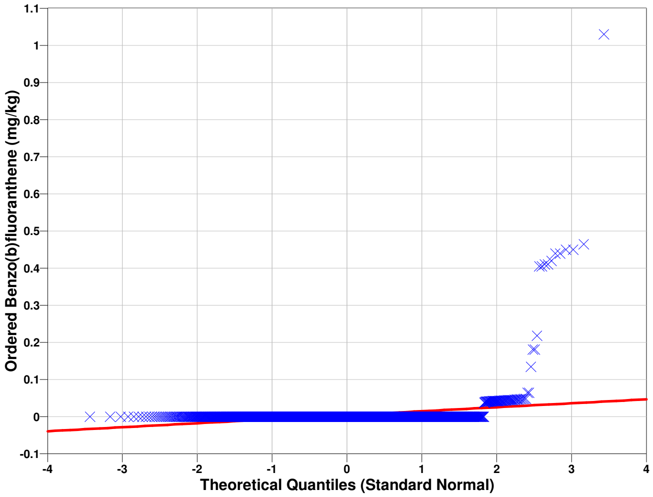
The Histogram is a plot of the fraction of the n observed data that fall within specified data "bins." A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25th and 75th percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The pth quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Benzo(b)fluoranthene
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Lilliefors test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Lilliefors Test Statistic	0.5084
Lilliefors 5% Critical Value	0.01936

The calculated Lilliefors test statistic exceeds the 5% Lilliefors critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.005438
95% Non-Parametric (Chebyshev) UCL	0.007726

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.007726) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=2094 data,
AL is the action level or threshold (0.1476),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=2093 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-170.23	1.6456	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
2080	1085	Reject

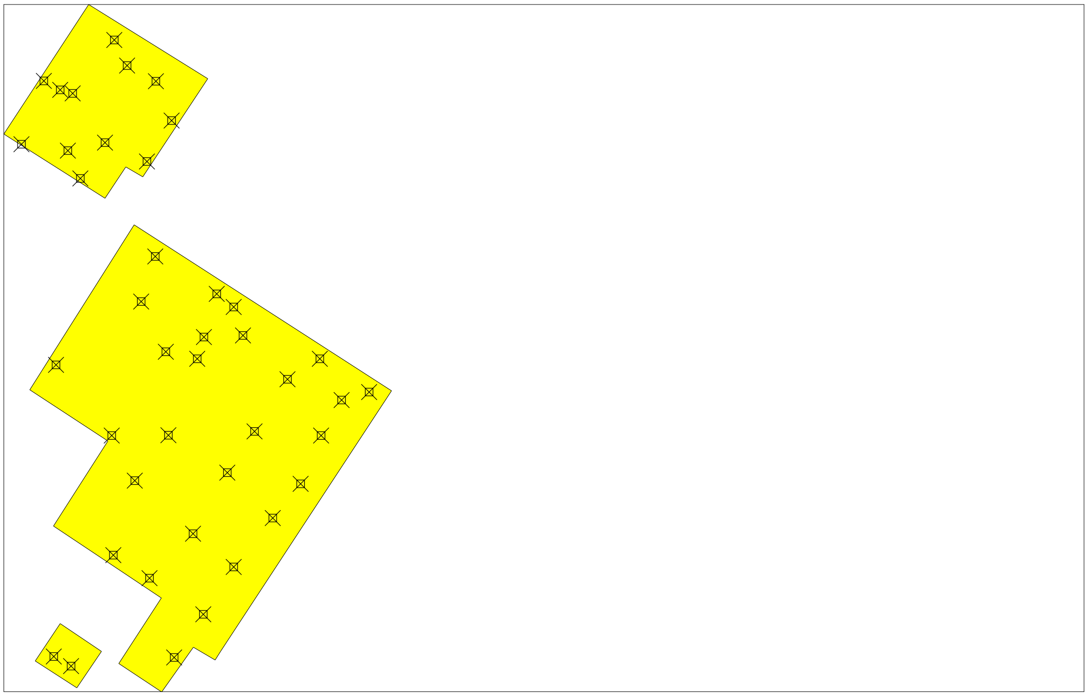
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	11.3	Manual	T
679301.1600	3083254.0340	J-12S	2.3	Manual	T
679171.2970	3083289.7960	J-04S	0.92	Manual	T
679155.0740	3083294.6960	J-03S	1.25	Manual	T
679133.4290	3083306.3130	J-01S	2.8	Manual	T
679104.2450	3083223.2620	J-02S	0.9	Manual	T
679164.8060	3083214.7100	J-06S	0.505	Manual	T
679181.2750	3083178.2880	J-08S	3.6	Manual	T
679213.7730	3083224.9730	J-09S	4.5	Manual	T
679280.5440	3083305.6810	J-10S	2.9	Manual	T
679242.7260	3083326.5280	J-07S	13.3	Manual	T
679225.8560	3083359.9740	J-05S	0.77	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	7.4	Manual	T
679335.0020	3082941.1720	J-21S	8.3	Manual	T
679343.5810	3082969.5980	J-19S	3.8	Manual	T
679394.8070	3082971.8300	J-24S	1.6	Manual	T
679382.8640	3083009.1130	J-20S	4	Manual	T
679293.5600	3082950.4980	J-17S	9	Manual	T
679360.5700	3083026.4980	J-18S	2.8	Manual	T

679297.0010	3082840.6970	J-23S	6.9	Manual	T
679252.7130	3082781.0290	J-22S	5.15	Manual	T
679222.6340	3082840.1720	J-16S	4	Manual	T
679279.6830	3083075.4290	J-14S	5.1	Manual	T
679149.4920	3082933.0980	J-13S	1.4	Manual	T
679272.0040	3082652.6750	J-28S	3.9	Manual	T
679224.5850	3082683.1400	J-26S	1.714	Manual	T
679146.6460	3082549.7640	J-25S	1.45	Manual	T
679169.0760	3082537.3510	J-27S	4.2	Manual	T
679342.7410	3082605.3190	J-35S	3.6	Manual	T
679304.6530	3082548.6880	J-34S	14.9	Manual	T
679382.8900	3082667.5270	J-36S	5.05	Manual	T
679329.4380	3082711.0960	J-29S	6.7	Manual	T
679374.4420	3082791.3300	J-30S	10.4	Manual	T
679453.4760	3082914.1150	J-32S	8.9	Manual	T
679410.1490	3082845.8460	J-31S	9.6	Manual	T
679560.6070	3082897.2580	J-41S	3.7	Manual	T
679495.8840	3082940.9730	J-33S	2	Manual	T
679524.3310	3082886.8990	J-40S	7.8	Manual	T
679497.3310	3082840.3960	J-39S	4.5	Manual	T
679470.3570	3082776.7350	J-38S	6.4	Manual	T
679433.9450	3082731.6820	J-37S	4.9	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

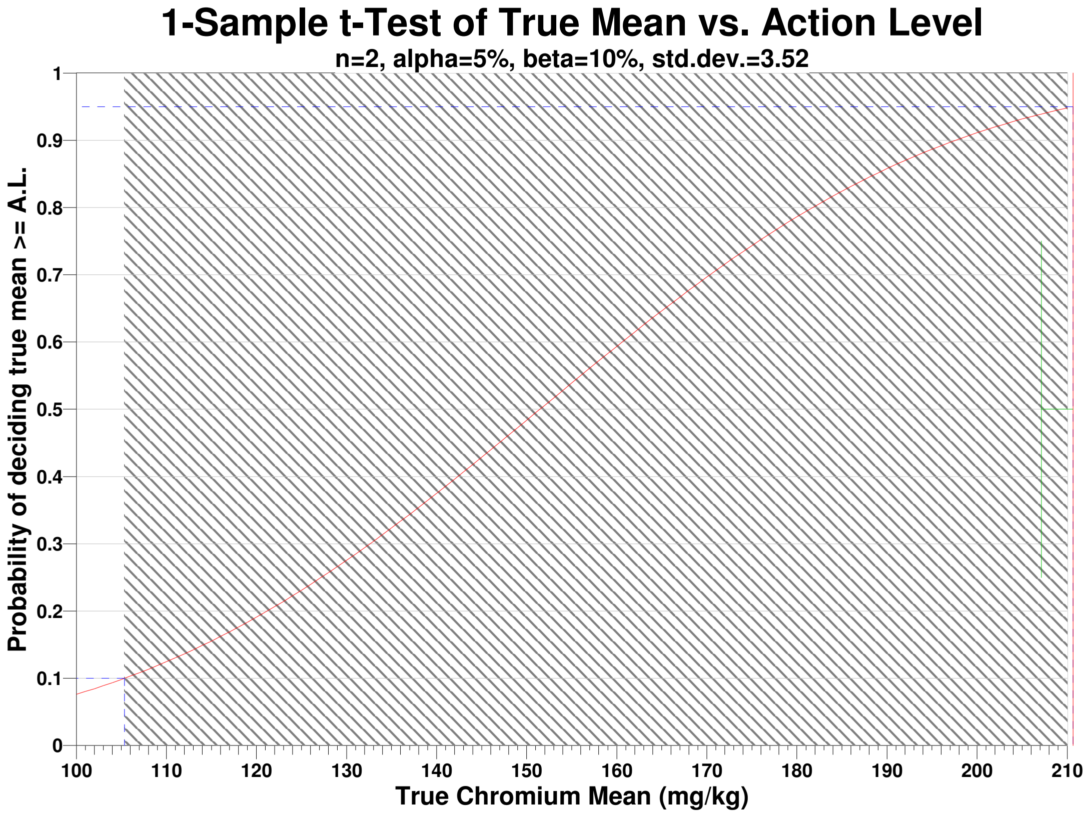
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Chromium	2	3.52 mg/kg	105.338 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=210.675		**α=5**		**α=10**		**α=15**	
		s=7.04	**s=3.52**	**s=7.04**	**s=3.52**	**s=7.04**	**s=3.52**
LBGR=90	**β=5**	3	2	2	2	2	1
	β=10	3	2	2	2	2	1
	β=15	3	2	2	1	2	1
LBGR=80	**β=5**	2	2	2	1	1	1
	β=10	2	2	2	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	**β=5**	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Chromium
The following data points were entered by the user for analysis.

Chromium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.505	0.77	0.9	0.92	1.25	1.4	1.45	1.6	1.714	2
10	2.3	2.8	2.8	2.9	3.6	3.6	3.7	3.8	3.9	4
20	4	4.2	4.5	4.5	4.9	5.05	5.1	5.15	6.4	6.7
30	6.9	7.4	7.8	8.3	8.9	9	9.6	10.4	11.3	13.3
40	14.9									

SUMMARY STATISTICS for Chromium	
n	41
Min	0.505
Max	14.9
Range	14.395
Mean	4.9807
Median	4
Variance	12.41
StdDev	3.5228
Std Error	0.55018
Skewness	1.0052

Interquartile Range				5				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.505	0.783	0.986	2.15	4	7.15	10.24	13.1	14.9

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Chromium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	2.816	3.05	No

None of the test statistics exceeded the corresponding critical values, therefore none of the 1 tests are significant and we conclude that at the 5% significance level there are no outliers in the data.

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.9268
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

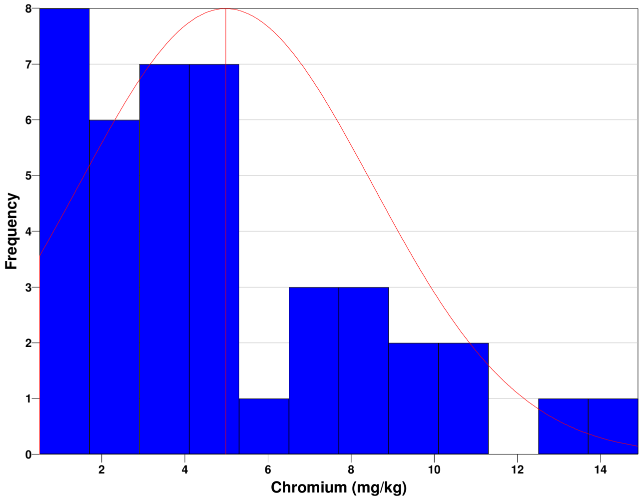
Data Plots for Chromium
Graphical displays of the data are shown below.

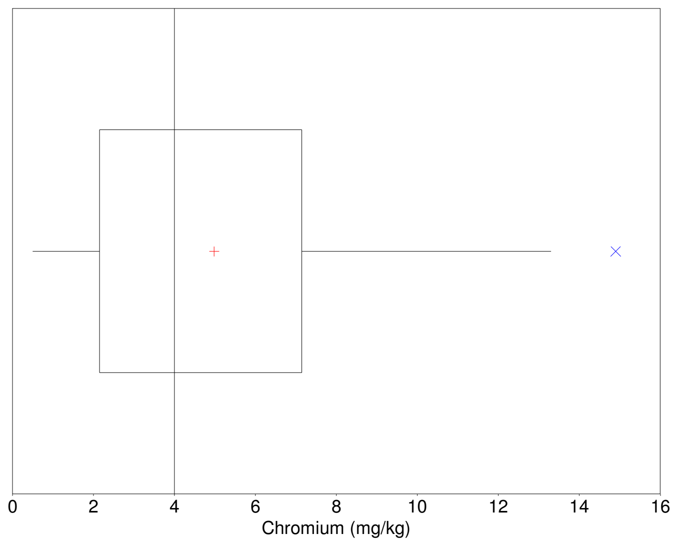
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

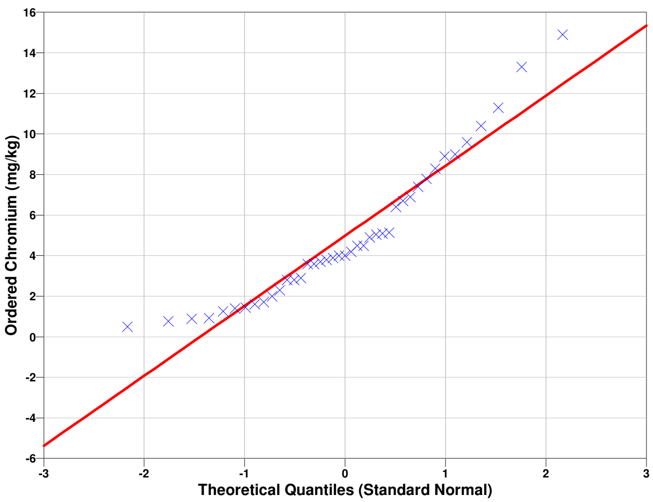
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the

distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Chromium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.9119
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	5.907
95% Non-Parametric (Chebyshev) UCL	7.379

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (7.379) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (210.675),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-373.87	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

Random sampling locations for comparing a mean with a fixed threshold (parametric)

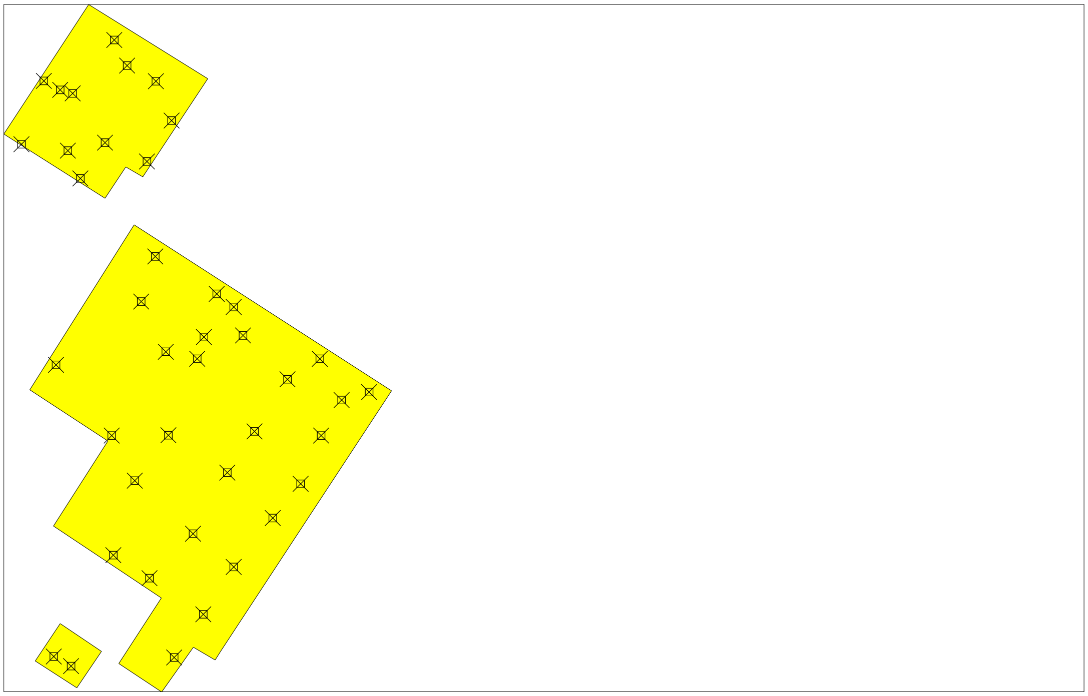
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	11.3	Manual	T
679301.1600	3083254.0340	J-12S	2.3	Manual	T
679171.2970	3083289.7960	J-04S	0.92	Manual	T
679155.0740	3083294.6960	J-03S	1.25	Manual	T
679133.4290	3083306.3130	J-01S	2.8	Manual	T
679104.2450	3083223.2620	J-02S	0.9	Manual	T
679164.8060	3083214.7100	J-06S	0.505	Manual	T
679181.2750	3083178.2880	J-08S	3.6	Manual	T
679213.7730	3083224.9730	J-09S	4.5	Manual	T
679280.5440	3083305.6810	J-10S	2.9	Manual	T
679242.7260	3083326.5280	J-07S	13.3	Manual	T
679225.8560	3083359.9740	J-05S	0.77	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	7.4	Manual	T
679335.0020	3082941.1720	J-21S	8.3	Manual	T
679343.5810	3082969.5980	J-19S	3.8	Manual	T
679394.8070	3082971.8300	J-24S	1.6	Manual	T
679382.8640	3083009.1130	J-20S	4	Manual	T
679293.5600	3082950.4980	J-17S	9	Manual	T
679360.5700	3083026.4980	J-18S	2.8	Manual	T

679297.0010	3082840.6970	J-23S	6.9	Manual	T
679252.7130	3082781.0290	J-22S	5.15	Manual	T
679222.6340	3082840.1720	J-16S	4	Manual	T
679279.6830	3083075.4290	J-14S	5.1	Manual	T
679149.4920	3082933.0980	J-13S	1.4	Manual	T
679272.0040	3082652.6750	J-28S	3.9	Manual	T
679224.5850	3082683.1400	J-26S	1.714	Manual	T
679146.6460	3082549.7640	J-25S	1.45	Manual	T
679169.0760	3082537.3510	J-27S	4.2	Manual	T
679342.7410	3082605.3190	J-35S	3.6	Manual	T
679304.6530	3082548.6880	J-34S	14.9	Manual	T
679382.8900	3082667.5270	J-36S	5.05	Manual	T
679329.4380	3082711.0960	J-29S	6.7	Manual	T
679374.4420	3082791.3300	J-30S	10.4	Manual	T
679453.4760	3082914.1150	J-32S	8.9	Manual	T
679410.1490	3082845.8460	J-31S	9.6	Manual	T
679560.6070	3082897.2580	J-41S	3.7	Manual	T
679495.8840	3082940.9730	J-33S	2	Manual	T
679524.3310	3082886.8990	J-40S	7.8	Manual	T
679497.3310	3082840.3960	J-39S	4.5	Manual	T
679470.3570	3082776.7350	J-38S	6.4	Manual	T
679433.9450	3082731.6820	J-37S	4.9	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

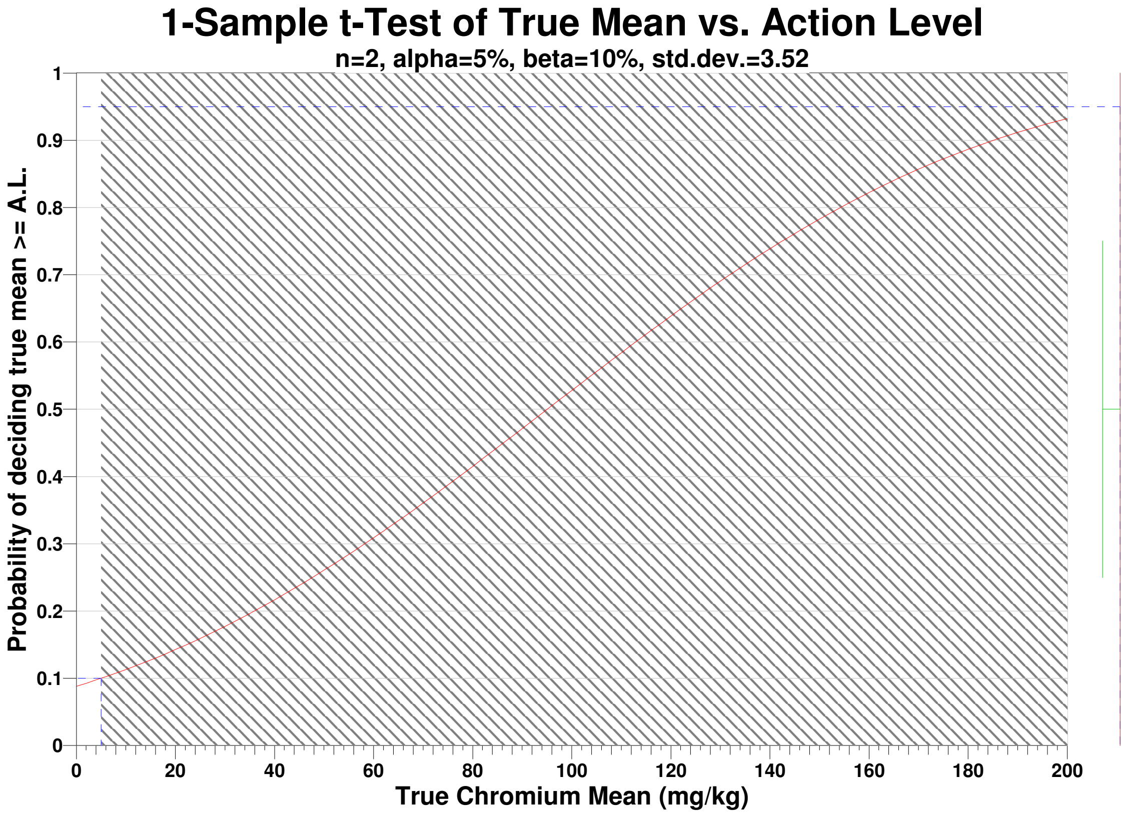
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Chromium	2	3.52 mg/kg	205.69 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=210.675		α=5		α=10		α=15	
		s=7.04	s=3.52	s=7.04	s=3.52	s=7.04	s=3.52
LBGR=90	β=5	3	2	2	2	2	1
	β=10	3	2	2	2	2	1
	β=15	3	2	2	1	2	1
LBGR=80	β=5	2	2	2	1	1	1
	β=10	2	2	2	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	β=5	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Chromium
The following data points were entered by the user for analysis.

Chromium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.505	0.77	0.9	0.92	1.25	1.4	1.45	1.6	1.714	2
10	2.3	2.8	2.8	2.9	3.6	3.6	3.7	3.8	3.9	4
20	4	4.2	4.5	4.5	4.9	5.05	5.1	5.15	6.4	6.7
30	6.9	7.4	7.8	8.3	8.9	9	9.6	10.4	11.3	13.3
40	14.9									

SUMMARY STATISTICS for Chromium	
n	41
Min	0.505
Max	14.9
Range	14.395
Mean	4.9807
Median	4
Variance	12.41
StdDev	3.5228
Std Error	0.55018
Skewness	1.0052

Interquartile Range				5				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.505	0.783	0.986	2.15	4	7.15	10.24	13.1	14.9

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Chromium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	2.816	3.05	No

None of the test statistics exceeded the corresponding critical values, therefore none of the 1 tests are significant and we conclude that at the 5% significance level there are no outliers in the data.

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.9268
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

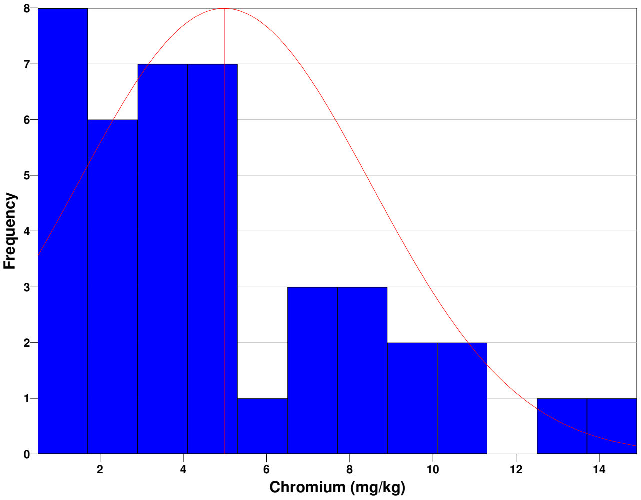
Data Plots for Chromium
Graphical displays of the data are shown below.

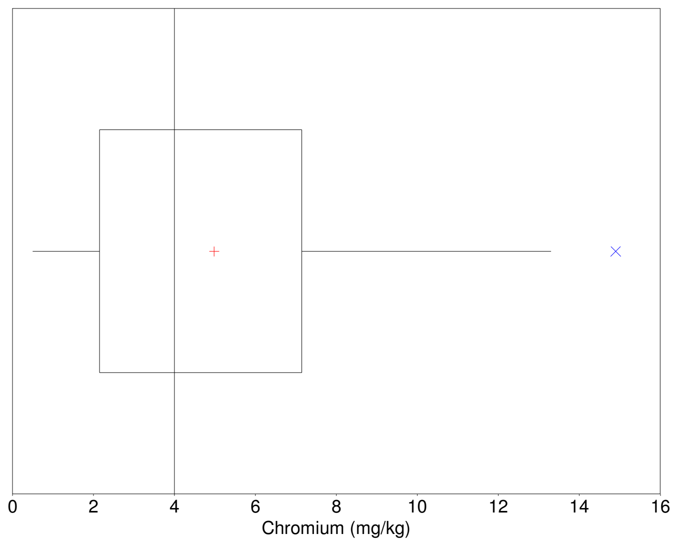
The Histogram is a plot of the fraction of the n observed data that fall within specified data "bins." A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

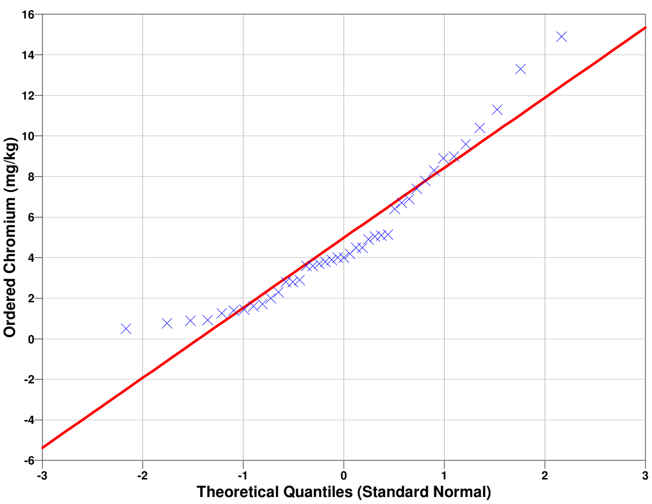
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the

distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Chromium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.9119
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	5.907
95% Non-Parametric (Chebyshev) UCL	7.379

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (7.379) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (210.675),
SE is the standard error = (standard deviation) / (square root of n).

This *t* was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-373.87	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

Random sampling locations for comparing a mean with a fixed threshold (parametric)

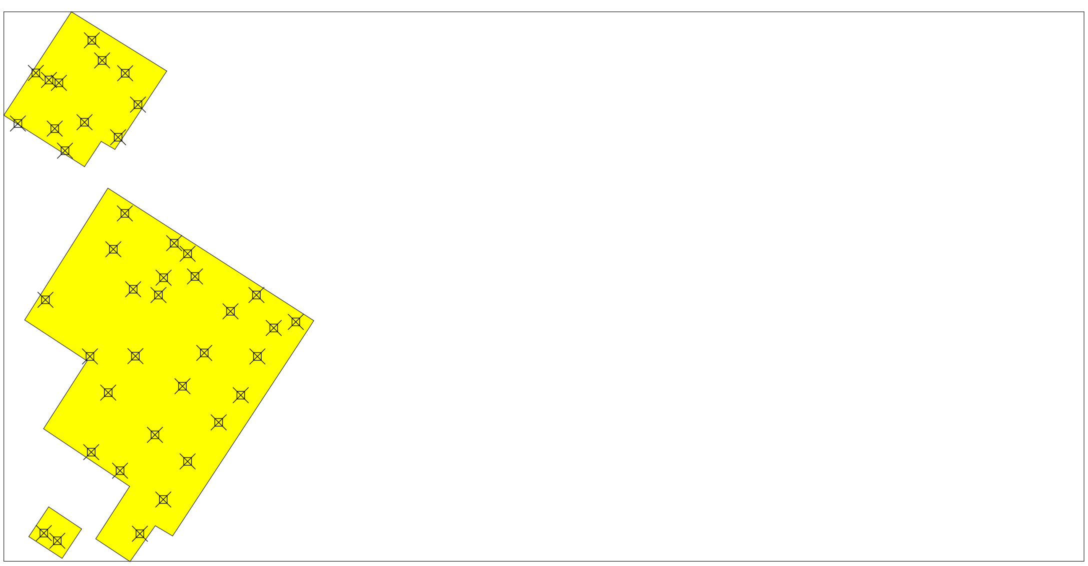
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1

X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.0037	Manual	T
679301.1600	3083254.0340	J-12S	0.0065	Manual	T
679171.2970	3083289.7960	J-04S	0.007	Manual	T
679155.0740	3083294.6960	J-03S	0.0096	Manual	T
679133.4290	3083306.3130	J-01S	0.0083	Manual	T
679104.2450	3083223.2620	J-02S	0.0069	Manual	T
679164.8060	3083214.7100	J-06S	0.00535	Manual	T
679181.2750	3083178.2880	J-08S	0.0014	Manual	T
679213.7730	3083224.9730	J-09S	0.013	Manual	T
679280.5440	3083305.6810	J-10S	0.0047	Manual	T
679242.7260	3083326.5280	J-07S	0.74	Manual	T
679225.8560	3083359.9740	J-05S	0.0014	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	0.011	Manual	T
679335.0020	3082941.1720	J-21S	0.016	Manual	T
679343.5810	3082969.5980	J-19S	0.049	Manual	T
679394.8070	3082971.8300	J-24S	0.026	Manual	T
679382.8640	3083009.1130	J-20S	0.013	Manual	T
679293.5600	3082950.4980	J-17S	0.024	Manual	T
679360.5700	3083026.4980	J-18S	0.0088	Manual	T
679297.0010	3082840.6970	J-23S	0.012	Manual	T
679252.7130	3082781.0290	J-22S	0.034	Manual	T
679222.6340	3082840.1720	J-16S	0.034	Manual	T
679279.6830	3083075.4290	J-14S	0.0054	Manual	T
679149.4920	3082933.0980	J-13S	0.000385	Manual	T
679272.0040	3082652.6750	J-28S	0.0093	Manual	T
679224.5850	3082683.1400	J-26S	0.0073	Manual	T
679146.6460	3082549.7640	J-25S	0.0095	Manual	T
679169.0760	3082537.3510	J-27S	0.016	Manual	T
679342.7410	3082605.3190	J-35S	0.0072	Manual	T
679304.6530	3082548.6880	J-34S	0.079	Manual	T
679382.8900	3082667.5270	J-36S	0.0031	Manual	T
679329.4380	3082711.0960	J-29S	0.014	Manual	T
679374.4420	3082791.3300	J-30S	0.024	Manual	T
679453.4760	3082914.1150	J-32S	0.007	Manual	T
679410.1490	3082845.8460	J-31S	0.0215	Manual	T
679560.6070	3082897.2580	J-41S	0.013	Manual	T
679495.8840	3082940.9730	J-33S	0.0076	Manual	T

679524.3310	3082886.8990	J-40S	0.0079	Manual	T
679497.3310	3082840.3960	J-39S	0.019	Manual	T
679470.3570	3082776.7350	J-38S	0.0054	Manual	T
679433.9450	3082731.6820	J-37S	0.0013	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

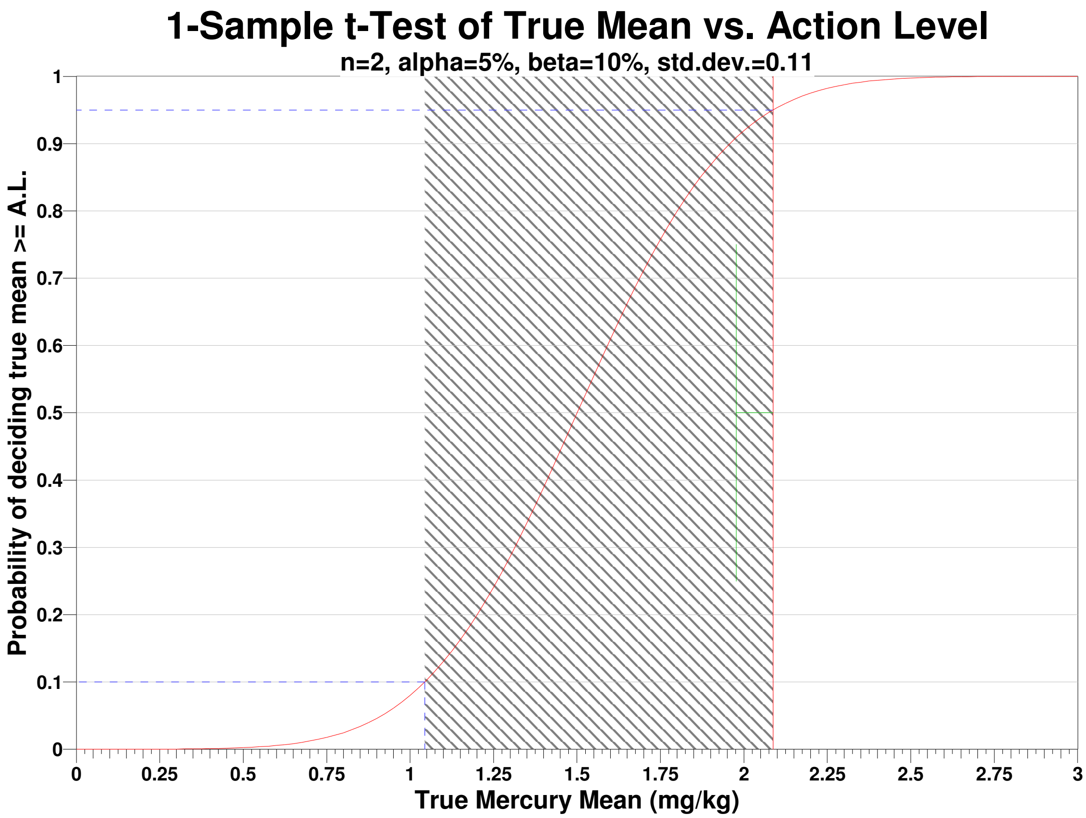
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	Δ	α	β	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Mercury	2	0.11 mg/kg	1.0436 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples			
AL=2.0872	α=5	α=10	α=15

		s=0.22	s=0.11	s=0.22	s=0.11	s=0.22	s=0.11	
LBGR=90	β=5	14	5	11	4	9	3
	β=10	11	4	9	3	7	3
	β=15	10	4	7	3	6	2
LBGR=80	β=5	5	3	4	2	3	2
	β=10	4	2	3	2	3	1
	β=15	4	2	3	2	2	1
LBGR=70	β=5	3	2	2	2	2	1
	β=10	3	2	2	2	2	1
	β=15	3	2	2	1	2	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Mercury
The following data points were entered by the user for analysis.

Mercury (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.000385	0.0013	0.0014	0.0014	0.0031	0.0037	0.0047	0.00535	0.0054	0.0054
10	0.0065	0.0069	0.007	0.007	0.0072	0.0073	0.0076	0.0079	0.0083	0.0088
20	0.0093	0.0095	0.0096	0.011	0.012	0.013	0.013	0.013	0.014	0.016
30	0.016	0.019	0.0215	0.024	0.024	0.026	0.034	0.034	0.049	0.079
40	0.74									

SUMMARY STATISTICS for Mercury	
n	41
Min	0.000385

Max				0.74				
Range				0.73962				
Mean				0.03155				
Median				0.0093				
Variance				0.013071				
StdDev				0.11433				
Std Error				0.017855				
Skewness				6.2475				
Interquartile Range				0.01155				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.000385	0.00131	0.00174	0.00595	0.0093	0.0175	0.034	0.076	0.74

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Mercury			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	6.197	3.05	Yes

The test statistic 6.197 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Mercury	
1	0.74

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.7183
Shapiro-Wilk 5% Critical Value	0.94

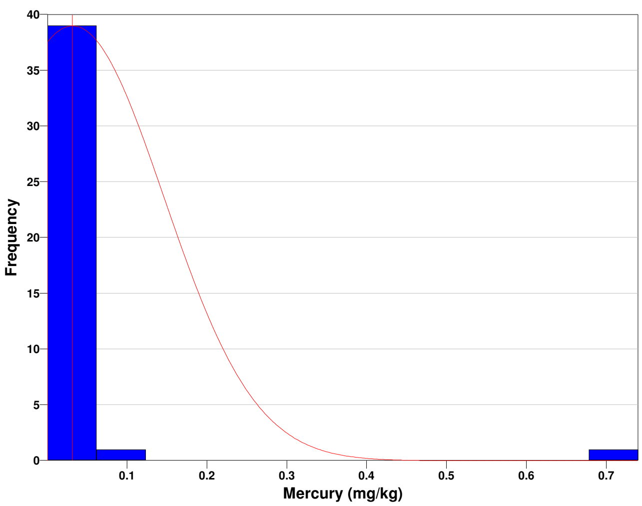
The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

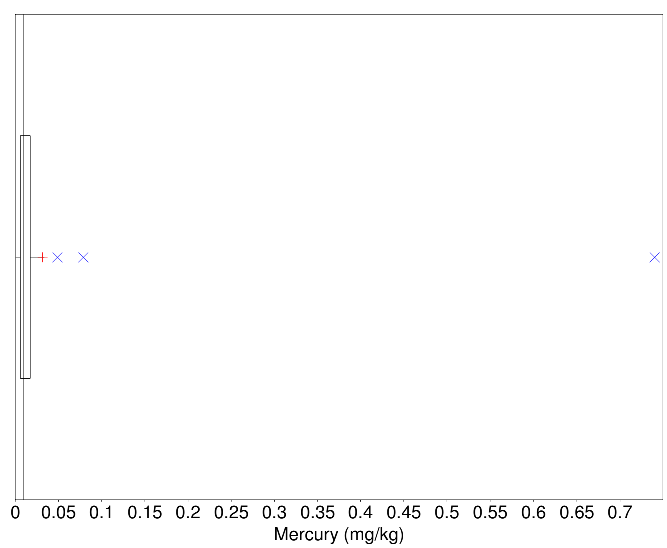
Data Plots for Mercury
Graphical displays of the data are shown below.

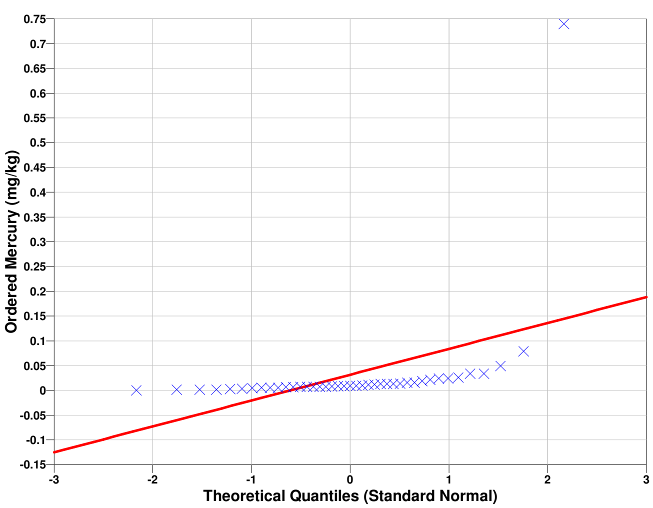
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through

2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Mercury
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.2381
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.06162
95% Non-Parametric (Chebyshev) UCL	0.1094

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.1094) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (2.0872),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-115.13	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

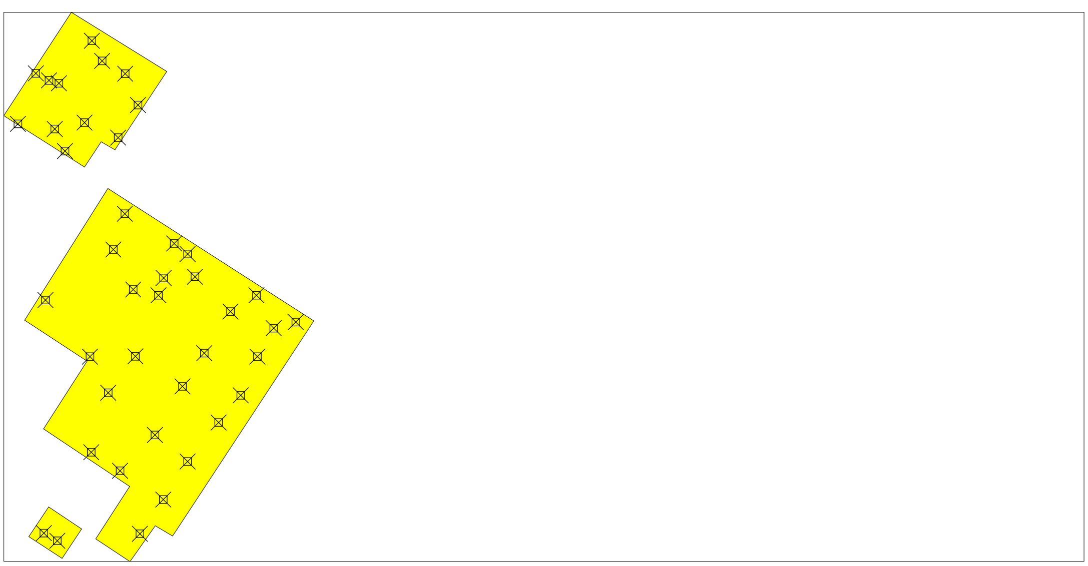
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1

X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	0.0037	Manual	T
679301.1600	3083254.0340	J-12S	0.0065	Manual	T
679171.2970	3083289.7960	J-04S	0.007	Manual	T
679155.0740	3083294.6960	J-03S	0.0096	Manual	T
679133.4290	3083306.3130	J-01S	0.0083	Manual	T
679104.2450	3083223.2620	J-02S	0.0069	Manual	T
679164.8060	3083214.7100	J-06S	0.00535	Manual	T
679181.2750	3083178.2880	J-08S	0.0014	Manual	T
679213.7730	3083224.9730	J-09S	0.013	Manual	T
679280.5440	3083305.6810	J-10S	0.0047	Manual	T
679242.7260	3083326.5280	J-07S	0.74	Manual	T
679225.8560	3083359.9740	J-05S	0.0014	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	0.011	Manual	T
679335.0020	3082941.1720	J-21S	0.016	Manual	T
679343.5810	3082969.5980	J-19S	0.049	Manual	T
679394.8070	3082971.8300	J-24S	0.026	Manual	T
679382.8640	3083009.1130	J-20S	0.013	Manual	T
679293.5600	3082950.4980	J-17S	0.024	Manual	T
679360.5700	3083026.4980	J-18S	0.0088	Manual	T
679297.0010	3082840.6970	J-23S	0.012	Manual	T
679252.7130	3082781.0290	J-22S	0.034	Manual	T
679222.6340	3082840.1720	J-16S	0.034	Manual	T
679279.6830	3083075.4290	J-14S	0.0054	Manual	T
679149.4920	3082933.0980	J-13S	0.000385	Manual	T
679272.0040	3082652.6750	J-28S	0.0093	Manual	T
679224.5850	3082683.1400	J-26S	0.0073	Manual	T
679146.6460	3082549.7640	J-25S	0.0095	Manual	T
679169.0760	3082537.3510	J-27S	0.016	Manual	T
679342.7410	3082605.3190	J-35S	0.0072	Manual	T
679304.6530	3082548.6880	J-34S	0.079	Manual	T
679382.8900	3082667.5270	J-36S	0.0031	Manual	T
679329.4380	3082711.0960	J-29S	0.014	Manual	T
679374.4420	3082791.3300	J-30S	0.024	Manual	T
679453.4760	3082914.1150	J-32S	0.007	Manual	T
679410.1490	3082845.8460	J-31S	0.0215	Manual	T
679560.6070	3082897.2580	J-41S	0.013	Manual	T
679495.8840	3082940.9730	J-33S	0.0076	Manual	T

679524.3310	3082886.8990	J-40S	0.0079	Manual	T
679497.3310	3082840.3960	J-39S	0.019	Manual	T
679470.3570	3082776.7350	J-38S	0.0054	Manual	T
679433.9450	3082731.6820	J-37S	0.0013	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

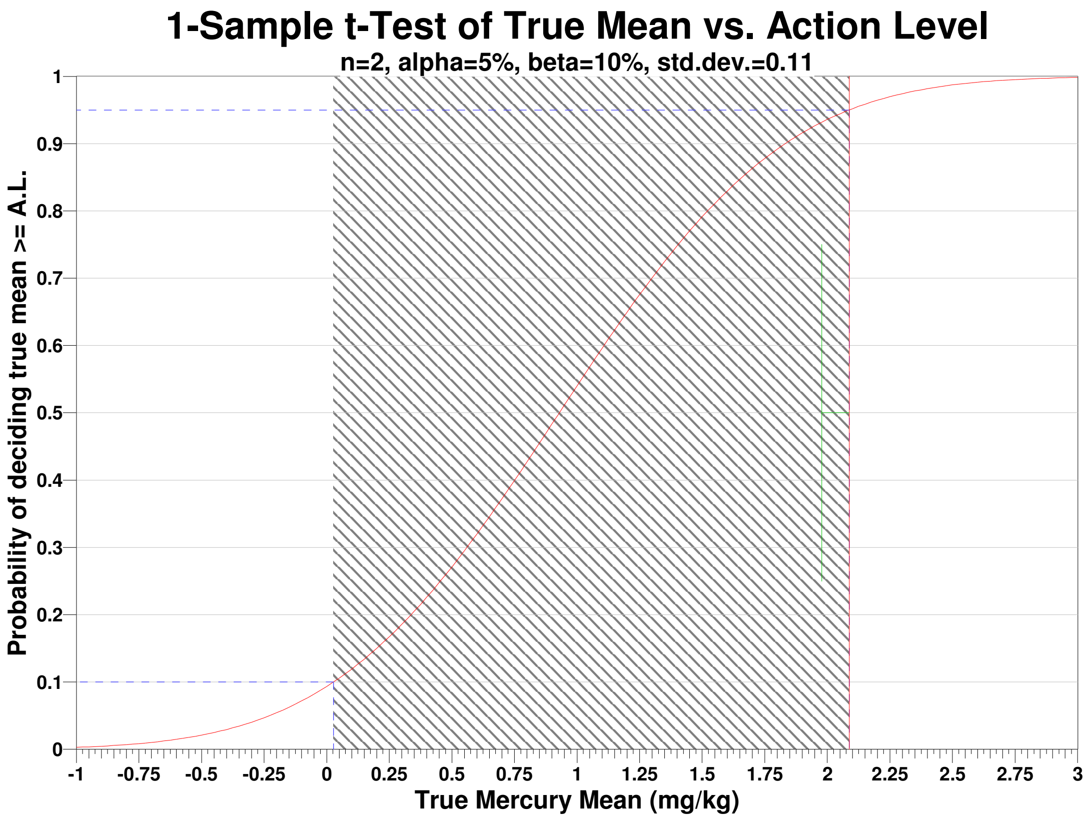
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Mercury	2	0.11 mg/kg	2.06 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples			
AL=2.0872	α=5	α=10	α=15

		s=0.22	s=0.11	s=0.22	s=0.11	s=0.22	s=0.11	
LBGR=90	β=5	14	5	11	4	9	3
	β=10	11	4	9	3	7	3
	β=15	10	4	7	3	6	2
LBGR=80	β=5	5	3	4	2	3	2
	β=10	4	2	3	2	3	1
	β=15	4	2	3	2	2	1
LBGR=70	β=5	3	2	2	2	2	1
	β=10	3	2	2	2	2	1
	β=15	3	2	2	1	2	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Mercury
The following data points were entered by the user for analysis.

Mercury (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.000385	0.0013	0.0014	0.0014	0.0031	0.0037	0.0047	0.00535	0.0054	0.0054
10	0.0065	0.0069	0.007	0.007	0.0072	0.0073	0.0076	0.0079	0.0083	0.0088
20	0.0093	0.0095	0.0096	0.011	0.012	0.013	0.013	0.013	0.014	0.016
30	0.016	0.019	0.0215	0.024	0.024	0.026	0.034	0.034	0.049	0.079
40	0.74									

SUMMARY STATISTICS for Mercury	
n	41
Min	0.000385

Max	0.74
Range	0.73962
Mean	0.03155
Median	0.0093
Variance	0.013071
StdDev	0.11433
Std Error	0.017855
Skewness	6.2475
Interquartile Range	0.01155

Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.000385	0.00131	0.00174	0.00595	0.0093	0.0175	0.034	0.076	0.74

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Mercury			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	6.197	3.05	Yes

The test statistic 6.197 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Mercury	
1	0.74

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.7183
Shapiro-Wilk 5% Critical Value	0.94

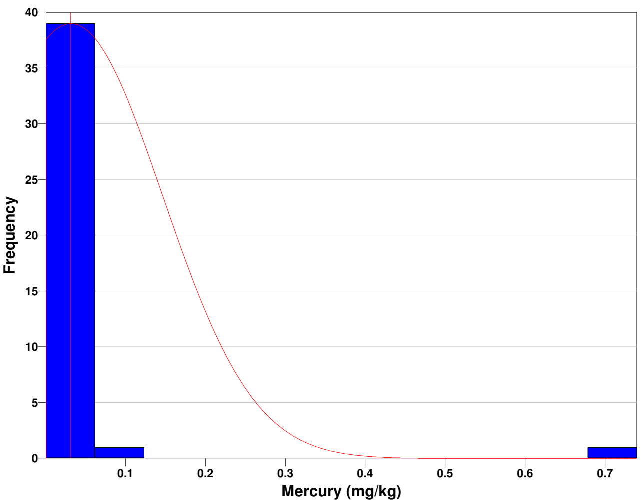
The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

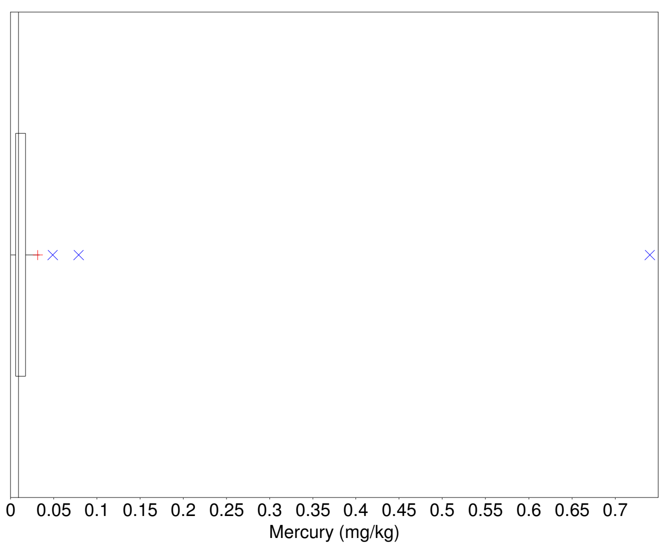
Data Plots for Mercury
Graphical displays of the data are shown below.

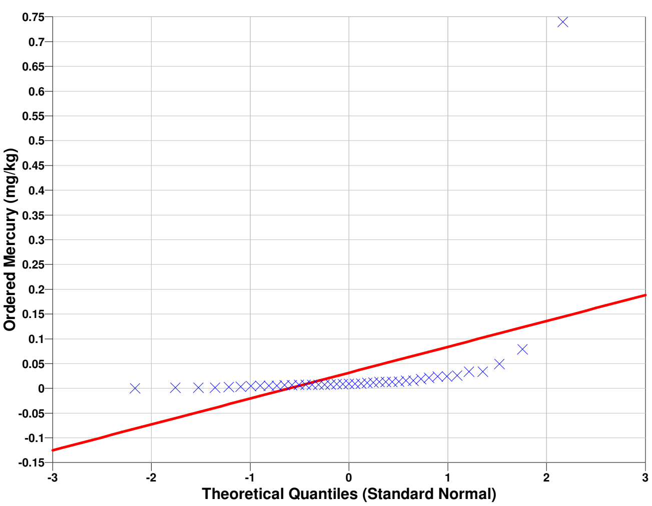
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through

2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Mercury
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.2381
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	0.06162
95% Non-Parametric (Chebyshev) UCL	0.1094

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (0.1094) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (2.0872),
SE is the standard error = (standard deviation) / (square root of n).

This *t* was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-115.13	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

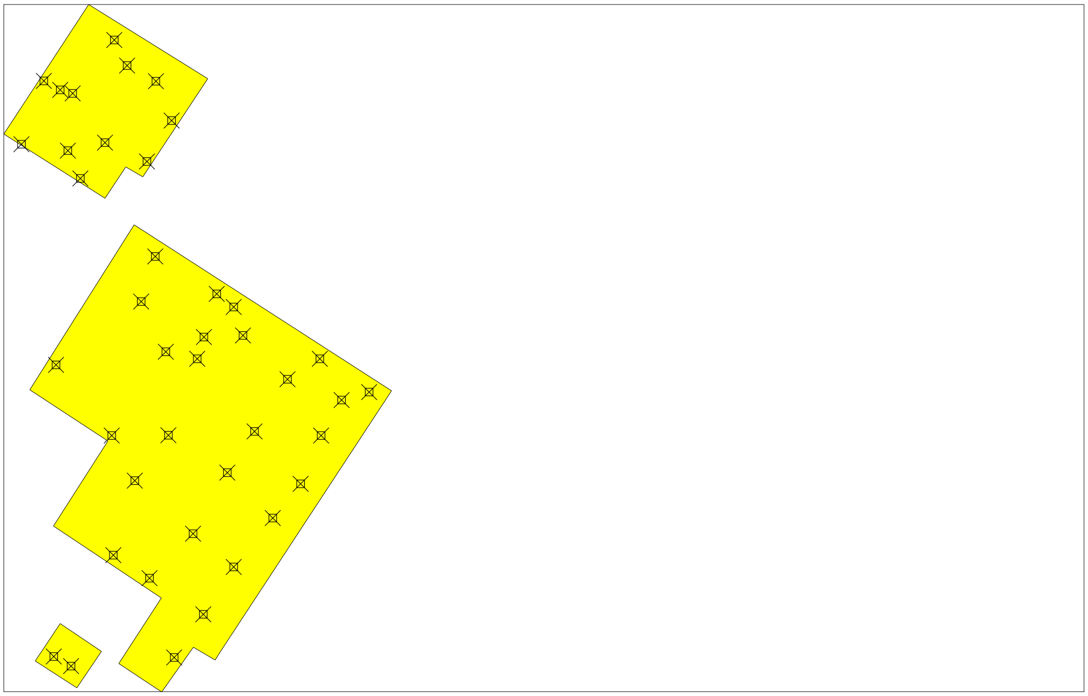
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	6	Manual	T
679301.1600	3083254.0340	J-12S	3.4	Manual	T
679171.2970	3083289.7960	J-04S	3.6	Manual	T
679155.0740	3083294.6960	J-03S	4.4	Manual	T
679133.4290	3083306.3130	J-01S	7	Manual	T
679104.2450	3083223.2620	J-02S	4.6	Manual	T
679164.8060	3083214.7100	J-06S	2.15	Manual	T
679181.2750	3083178.2880	J-08S	8.3	Manual	T
679213.7730	3083224.9730	J-09S	8.9	Manual	T
679280.5440	3083305.6810	J-10S	6.9	Manual	T
679242.7260	3083326.5280	J-07S	80.6	Manual	T
679225.8560	3083359.9740	J-05S	4.8	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	18.9	Manual	T
679335.0020	3082941.1720	J-21S	17.7	Manual	T
679343.5810	3082969.5980	J-19S	19.75	Manual	T
679394.8070	3082971.8300	J-24S	7.2	Manual	T
679382.8640	3083009.1130	J-20S	17.1	Manual	T
679293.5600	3082950.4980	J-17S	16.1	Manual	T
679360.5700	3083026.4980	J-18S	80.7	Manual	T

679297.0010	3082840.6970	J-23S	9.9	Manual	T
679252.7130	3082781.0290	J-22S	22.45	Manual	T
679222.6340	3082840.1720	J-16S	9.8	Manual	T
679279.6830	3083075.4290	J-14S	55.8	Manual	T
679149.4920	3082933.0980	J-13S	2.6	Manual	T
679272.0040	3082652.6750	J-28S	6.1	Manual	T
679224.5850	3082683.1400	J-26S	2.021	Manual	T
679146.6460	3082549.7640	J-25S	2.45	Manual	T
679169.0760	3082537.3510	J-27S	3.7	Manual	T
679342.7410	3082605.3190	J-35S	5.1	Manual	T
679304.6530	3082548.6880	J-34S	10.4	Manual	T
679382.8900	3082667.5270	J-36S	18.8	Manual	T
679329.4380	3082711.0960	J-29S	12.7	Manual	T
679374.4420	3082791.3300	J-30S	23.8	Manual	T
679453.4760	3082914.1150	J-32S	11.6	Manual	T
679410.1490	3082845.8460	J-31S	14.75	Manual	T
679560.6070	3082897.2580	J-41S	9.2	Manual	T
679495.8840	3082940.9730	J-33S	5.4	Manual	T
679524.3310	3082886.8990	J-40S	20.9	Manual	T
679497.3310	3082840.3960	J-39S	6.9	Manual	T
679470.3570	3082776.7350	J-38S	8	Manual	T
679433.9450	3082731.6820	J-37S	6.7	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

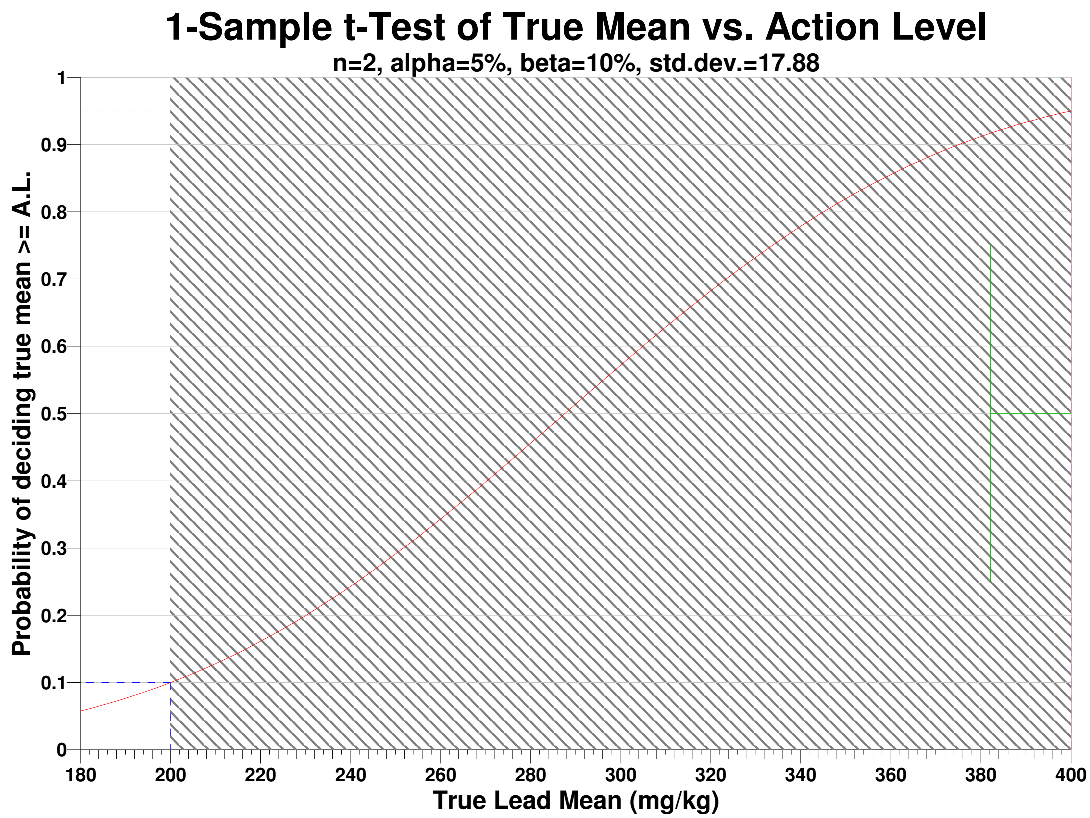
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Lead	2	17.88 mg/kg	200 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=400		α=5		α=10		α=15	
		s=35.76	s=17.88	s=35.76	s=17.88	s=35.76	s=17.88
LBGR=90	β=5	11	4	8	3	7	2
	β=10	9	4	7	3	5	2
	β=15	8	3	6	2	4	2
LBGR=80	β=5	4	2	3	2	2	1
	β=10	4	2	3	2	2	1
	β=15	3	2	2	2	2	1
LBGR=70	β=5	3	2	2	2	2	1

	β=10	3	2	2	1	2	1
	β=15	2	2	2	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Lead
The following data points were entered by the user for analysis.

Lead (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	2.021	2.15	2.45	2.6	3.4	3.6	3.7	4.4	4.6	4.8
10	5.1	5.4	6	6.1	6.7	6.9	6.9	7	7.2	8
20	8.3	8.9	9.2	9.8	9.9	10.4	11.6	12.7	14.75	16.1
30	17.1	17.7	18.8	18.9	19.75	20.9	22.45	23.8	55.8	80.6
40	80.7									

SUMMARY STATISTICS for Lead	
n	41
Min	2.021
Max	80.7
Range	78.679
Mean	14.321
Median	8.3
Variance	319.78
StdDev	17.882
Std Error	2.7928
Skewness	2.9321

Interquartile Range				12.45				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
2.021	2.18	2.76	4.95	8.3	17.4	23.53	78.12	80.7

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Lead			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.712	3.05	Yes

The test statistic 3.712 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Lead	
1	80.7

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.6201
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

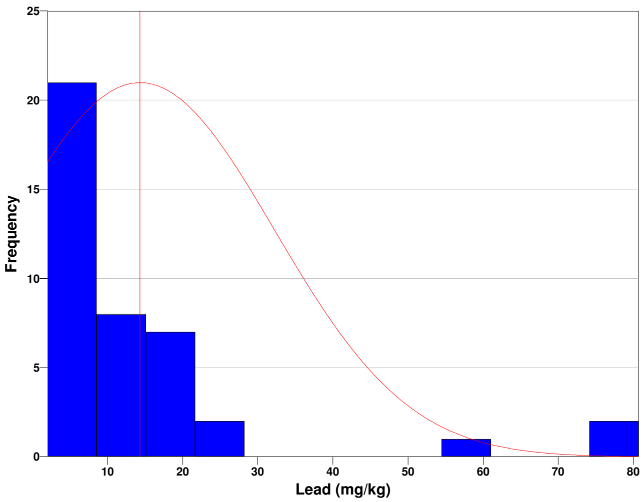
Data Plots for Lead
Graphical displays of the data are shown below.

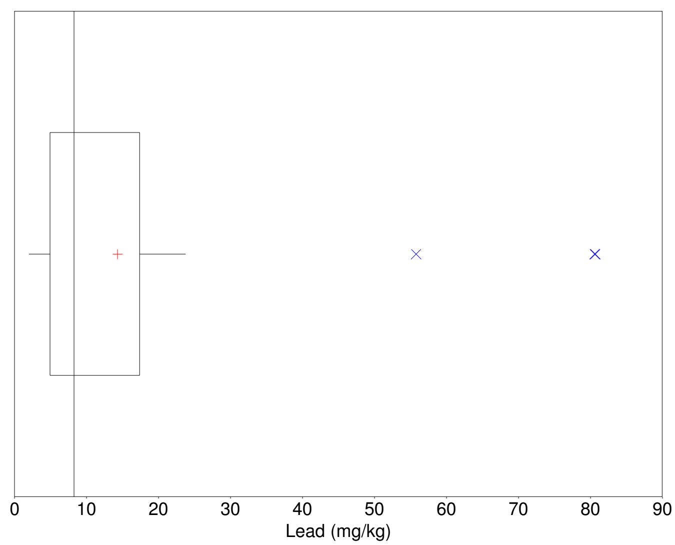
The Histogram is a plot of the fraction of the n observed data that fall within specified data "bins." A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

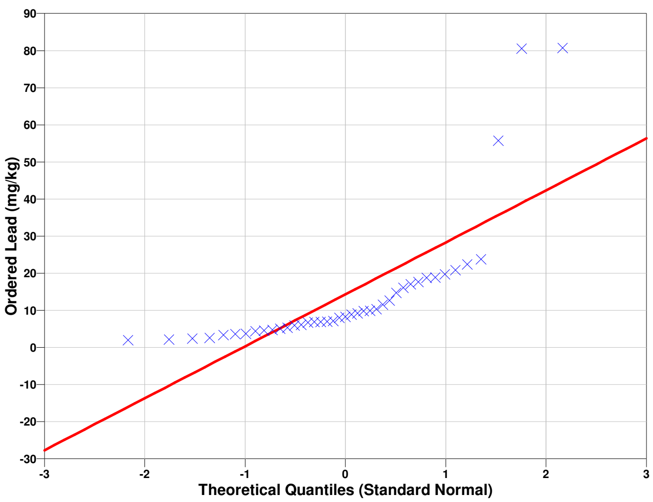
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Lead
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.5986
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	19.02
95% Non-Parametric (Chebyshev) UCL	26.49

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (26.49) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (400),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-138.1	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

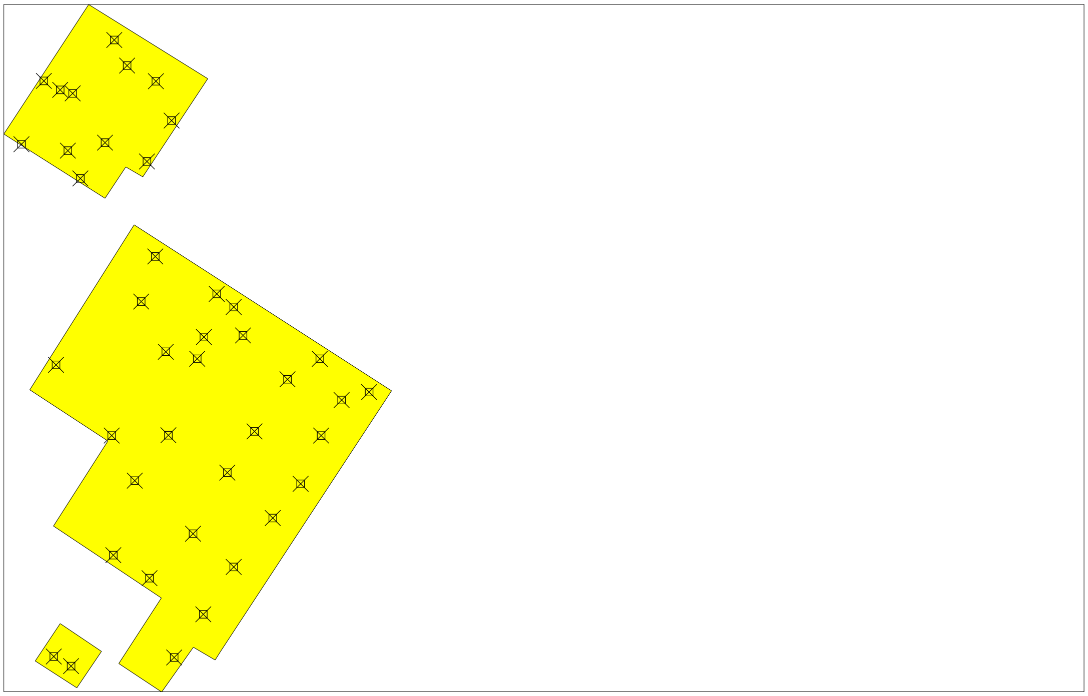
Random sampling locations for comparing a mean with a fixed threshold (parametric)

Summary
This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	6	Manual	T
679301.1600	3083254.0340	J-12S	3.4	Manual	T
679171.2970	3083289.7960	J-04S	3.6	Manual	T
679155.0740	3083294.6960	J-03S	4.4	Manual	T
679133.4290	3083306.3130	J-01S	7	Manual	T
679104.2450	3083223.2620	J-02S	4.6	Manual	T
679164.8060	3083214.7100	J-06S	2.15	Manual	T
679181.2750	3083178.2880	J-08S	8.3	Manual	T
679213.7730	3083224.9730	J-09S	8.9	Manual	T
679280.5440	3083305.6810	J-10S	6.9	Manual	T
679242.7260	3083326.5280	J-07S	80.6	Manual	T
679225.8560	3083359.9740	J-05S	4.8	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	18.9	Manual	T
679335.0020	3082941.1720	J-21S	17.7	Manual	T
679343.5810	3082969.5980	J-19S	19.75	Manual	T
679394.8070	3082971.8300	J-24S	7.2	Manual	T
679382.8640	3083009.1130	J-20S	17.1	Manual	T
679293.5600	3082950.4980	J-17S	16.1	Manual	T
679360.5700	3083026.4980	J-18S	80.7	Manual	T

679297.0010	3082840.6970	J-23S	9.9	Manual	T
679252.7130	3082781.0290	J-22S	22.45	Manual	T
679222.6340	3082840.1720	J-16S	9.8	Manual	T
679279.6830	3083075.4290	J-14S	55.8	Manual	T
679149.4920	3082933.0980	J-13S	2.6	Manual	T
679272.0040	3082652.6750	J-28S	6.1	Manual	T
679224.5850	3082683.1400	J-26S	2.021	Manual	T
679146.6460	3082549.7640	J-25S	2.45	Manual	T
679169.0760	3082537.3510	J-27S	3.7	Manual	T
679342.7410	3082605.3190	J-35S	5.1	Manual	T
679304.6530	3082548.6880	J-34S	10.4	Manual	T
679382.8900	3082667.5270	J-36S	18.8	Manual	T
679329.4380	3082711.0960	J-29S	12.7	Manual	T
679374.4420	3082791.3300	J-30S	23.8	Manual	T
679453.4760	3082914.1150	J-32S	11.6	Manual	T
679410.1490	3082845.8460	J-31S	14.75	Manual	T
679560.6070	3082897.2580	J-41S	9.2	Manual	T
679495.8840	3082940.9730	J-33S	5.4	Manual	T
679524.3310	3082886.8990	J-40S	20.9	Manual	T
679497.3310	3082840.3960	J-39S	6.9	Manual	T
679470.3570	3082776.7350	J-38S	8	Manual	T
679433.9450	3082731.6820	J-37S	6.7	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

- n is the number of samples,
- S is the estimated standard deviation of the measured values including analytical error,
- Δ is the width of the gray region,
- α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
- β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
- $Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
- $Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

The values of these inputs that result in the calculated number of sampling locations are:

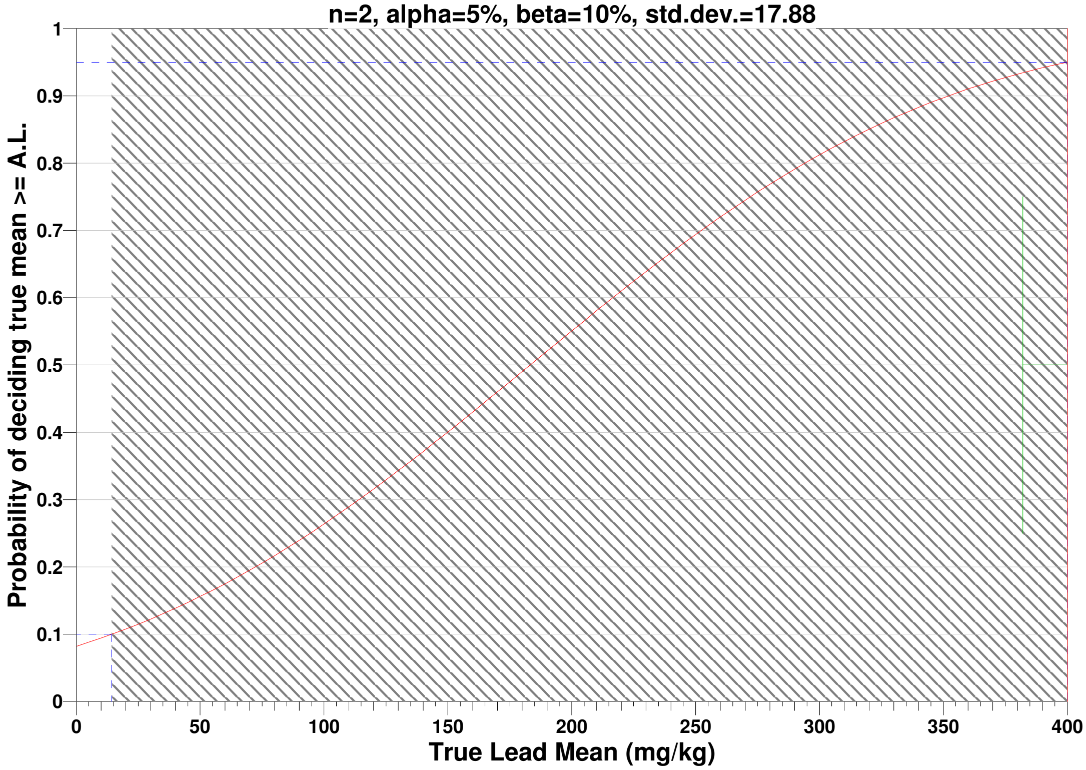
Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Lead	2	17.88 mg/kg	385.68 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.

1-Sample t-Test of True Mean vs. Action Level



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=400		α=5		α=10		α=15	
		s=35.76	s=17.88	s=35.76	s=17.88	s=35.76	s=17.88
LBGR=90	β=5	11	4	8	3	7	2
	β=10	9	4	7	3	5	2
	β=15	8	3	6	2	4	2
LBGR=80	β=5	4	2	3	2	2	1
	β=10	4	2	3	2	2	1
	β=15	3	2	2	2	2	1
LBGR=70	β=5	3	2	2	2	2	1

	β=10	3	2	2	1	2	1
	β=15	2	2	2	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Lead
The following data points were entered by the user for analysis.

Lead (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	2.021	2.15	2.45	2.6	3.4	3.6	3.7	4.4	4.6	4.8
10	5.1	5.4	6	6.1	6.7	6.9	6.9	7	7.2	8
20	8.3	8.9	9.2	9.8	9.9	10.4	11.6	12.7	14.75	16.1
30	17.1	17.7	18.8	18.9	19.75	20.9	22.45	23.8	55.8	80.6
40	80.7									

SUMMARY STATISTICS for Lead	
n	41
Min	2.021
Max	80.7
Range	78.679
Mean	14.321
Median	8.3
Variance	319.78
StdDev	17.882
Std Error	2.7928
Skewness	2.9321

Interquartile Range				12.45				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
2.021	2.18	2.76	4.95	8.3	17.4	23.53	78.12	80.7

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Lead			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.712	3.05	Yes

The test statistic 3.712 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Lead	
1	80.7

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.6201
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

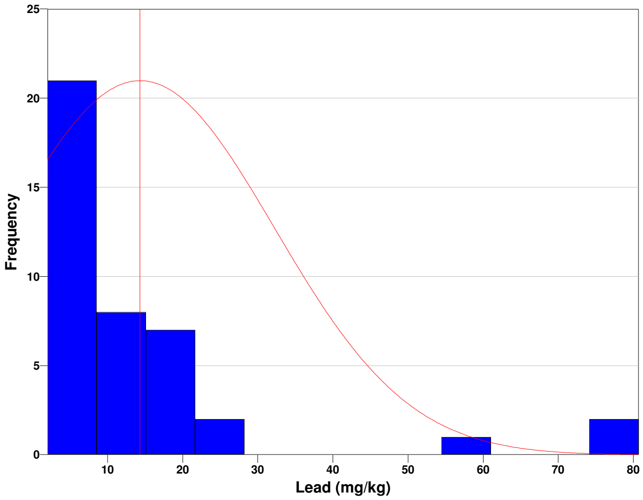
Data Plots for Lead
Graphical displays of the data are shown below.

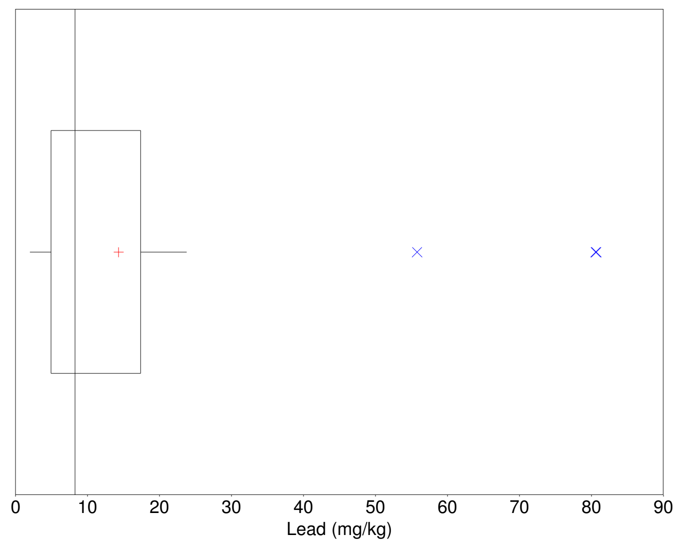
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

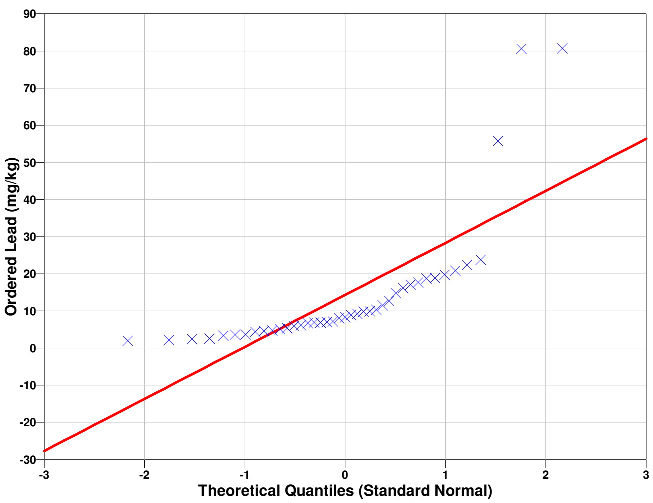
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Lead

A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.5986
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	19.02
95% Non-Parametric (Chebyshev) UCL	26.49

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (26.49) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (400),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-138.1	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

Random sampling locations for comparing a mean with a fixed threshold (parametric)

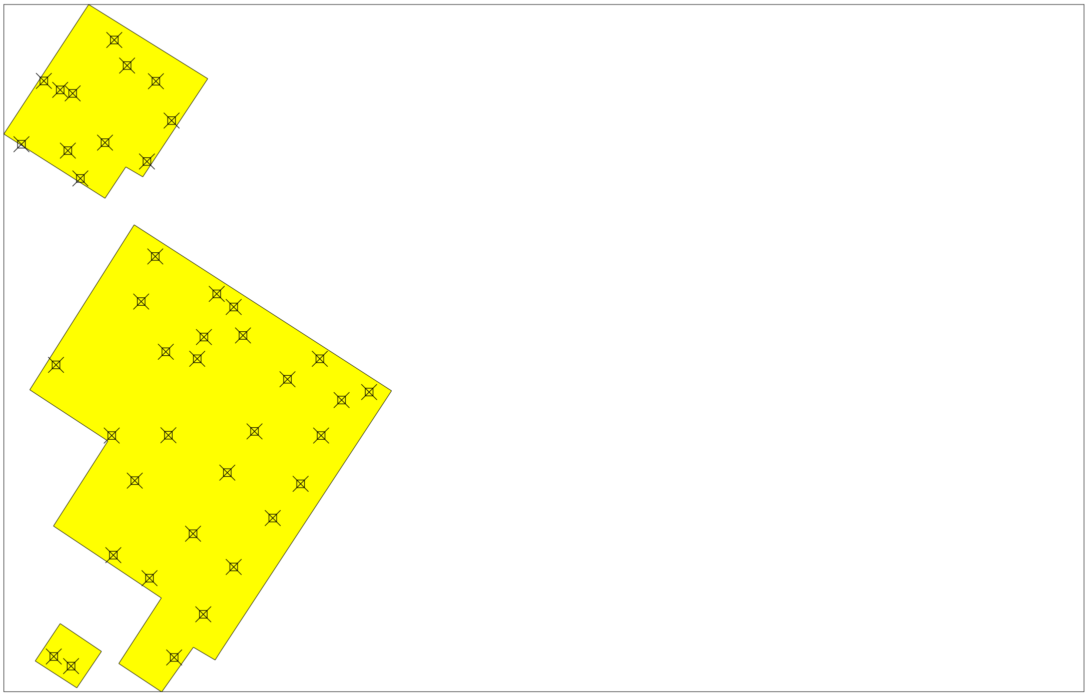
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	22.3	Manual	T
679301.1600	3083254.0340	J-12S	5.1	Manual	T
679171.2970	3083289.7960	J-04S	1.3	Manual	T
679155.0740	3083294.6960	J-03S	1.15	Manual	T
679133.4290	3083306.3130	J-01S	4.7	Manual	T
679104.2450	3083223.2620	J-02S	1.1	Manual	T
679164.8060	3083214.7100	J-06S	0.985	Manual	T
679181.2750	3083178.2880	J-08S	6.1	Manual	T
679213.7730	3083224.9730	J-09S	7.7	Manual	T
679280.5440	3083305.6810	J-10S	2.4	Manual	T
679242.7260	3083326.5280	J-07S	5.1	Manual	T
679225.8560	3083359.9740	J-05S	1.7	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	2.1	Manual	T
679335.0020	3082941.1720	J-21S	4.9	Manual	T
679343.5810	3082969.5980	J-19S	5.25	Manual	T
679394.8070	3082971.8300	J-24S	2.3	Manual	T
679382.8640	3083009.1130	J-20S	5.8	Manual	T
679293.5600	3082950.4980	J-17S	10.6	Manual	T
679360.5700	3083026.4980	J-18S	4.1	Manual	T

679297.0010	3082840.6970	J-23S	9.6	Manual	T
679252.7130	3082781.0290	J-22S	6.85	Manual	T
679222.6340	3082840.1720	J-16S	4.5	Manual	T
679279.6830	3083075.4290	J-14S	4.1	Manual	T
679149.4920	3082933.0980	J-13S	2.4	Manual	T
679272.0040	3082652.6750	J-28S	6.6	Manual	T
679224.5850	3082683.1400	J-26S	2.925	Manual	T
679146.6460	3082549.7640	J-25S	1.6	Manual	T
679169.0760	3082537.3510	J-27S	3.5	Manual	T
679342.7410	3082605.3190	J-35S	9.1	Manual	T
679304.6530	3082548.6880	J-34S	29.3	Manual	T
679382.8900	3082667.5270	J-36S	10.5	Manual	T
679329.4380	3082711.0960	J-29S	12.8	Manual	T
679374.4420	3082791.3300	J-30S	16.6	Manual	T
679453.4760	3082914.1150	J-32S	16.2	Manual	T
679410.1490	3082845.8460	J-31S	17.25	Manual	T
679560.6070	3082897.2580	J-41S	7	Manual	T
679495.8840	3082940.9730	J-33S	4.8	Manual	T
679524.3310	3082886.8990	J-40S	15.8	Manual	T
679497.3310	3082840.3960	J-39S	7.8	Manual	T
679470.3570	3082776.7350	J-38S	16	Manual	T
679433.9450	3082731.6820	J-37S	13.2	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5 Z_{1-\alpha}^2$$

where
n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

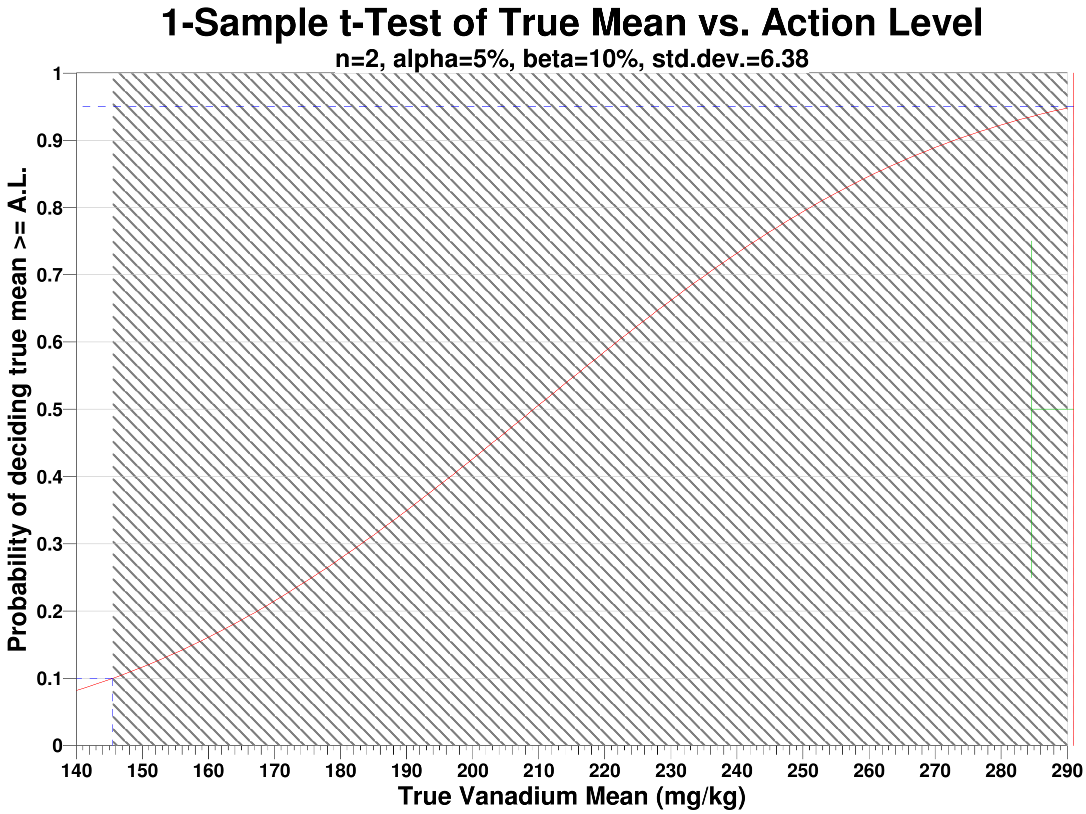
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	**n**	**Parameter**					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Vanadium	2	6.38 mg/kg	145.507 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=291.014		α=5		α=10		α=15	
		s=12.76	s=6.38	s=12.76	s=6.38	s=12.76	s=6.38
LBGR=90	β=5	4	2	3	2	2	1
	β=10	3	2	3	2	2	1
	β=15	3	2	2	2	2	1
LBGR=80	β=5	2	2	2	1	1	1
	β=10	2	2	2	1	1	1
	β=15	2	2	2	1	1	1
LBGR=70	β=5	2	2	2	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Vanadium
The following data points were entered by the user for analysis.

Vanadium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.985	1.1	1.15	1.3	1.6	1.7	2.1	2.3	2.4	2.4
10	2.925	3.5	4.1	4.1	4.5	4.7	4.8	4.9	5.1	5.1
20	5.25	5.8	6.1	6.6	6.85	7	7.7	7.8	9.1	9.6
30	10.5	10.6	12.8	13.2	15.8	16	16.2	16.6	17.25	22.3
40	29.3									

SUMMARY STATISTICS for Vanadium	
n	41
Min	0.985
Max	29.3
Range	28.315
Mean	7.6368
Median	5.25
Variance	40.68
StdDev	6.3781
Std Error	0.99609
Skewness	1.4681

Interquartile Range				7.8875				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.985	1.105	1.36	2.662	5.25	10.55	16.52	21.79	29.3

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Vanadium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.397	3.05	Yes

The test statistic 3.397 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Vanadium	
1	29.3

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.8816
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

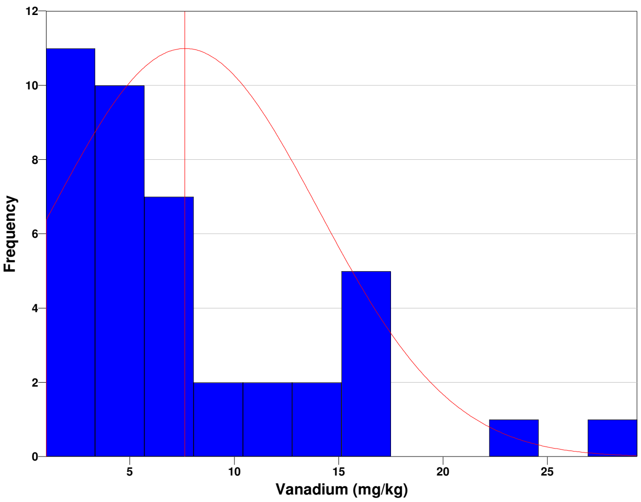
Data Plots for Vanadium
Graphical displays of the data are shown below.

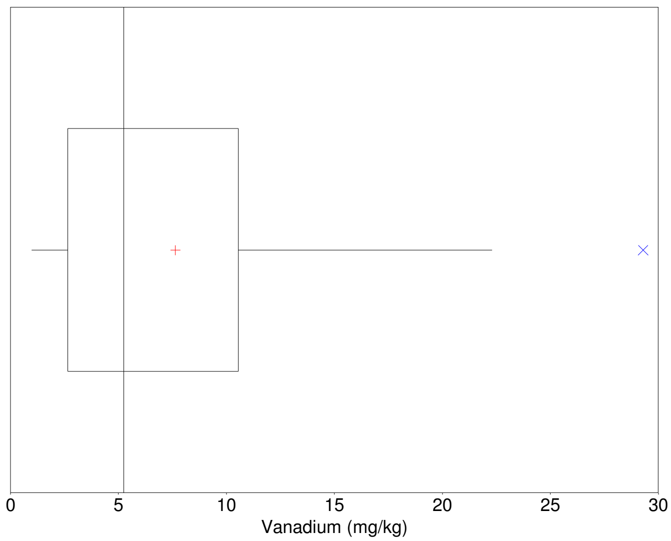
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

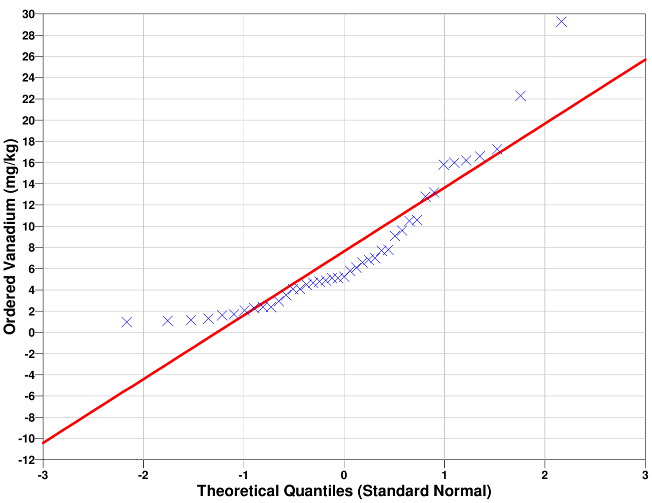
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Vanadium
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.8546
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	9.314
95% Non-Parametric (Chebyshev) UCL	11.98

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (11.98) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (291.014),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-284.49	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

Random sampling locations for comparing a mean with a fixed threshold (parametric)

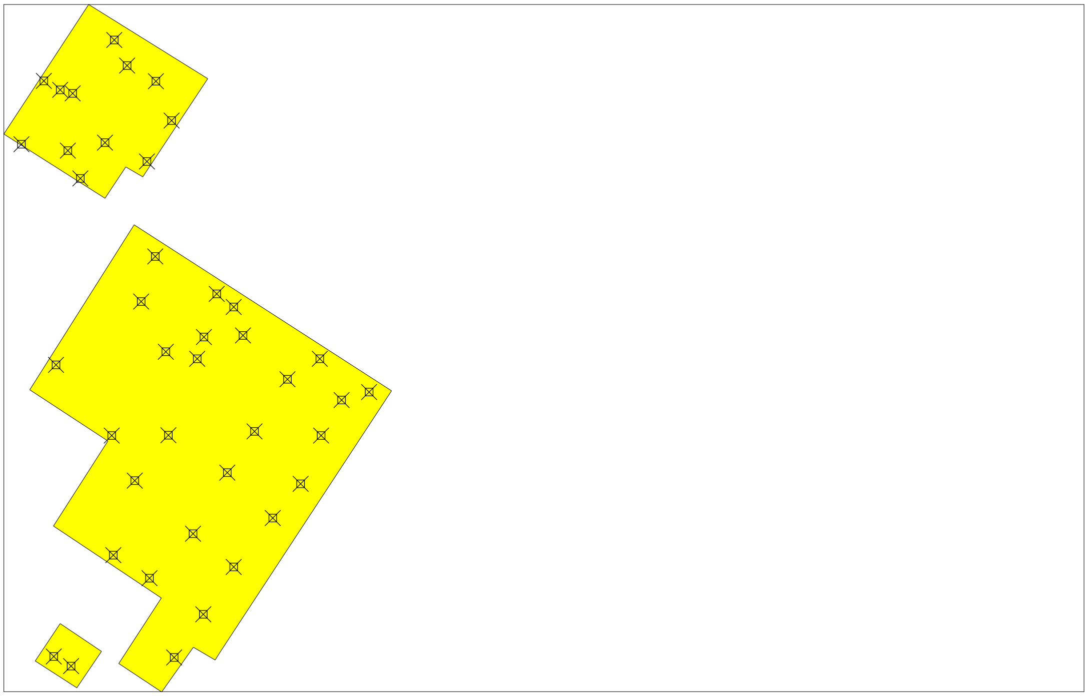
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	22.3	Manual	T
679301.1600	3083254.0340	J-12S	5.1	Manual	T
679171.2970	3083289.7960	J-04S	1.3	Manual	T
679155.0740	3083294.6960	J-03S	1.15	Manual	T
679133.4290	3083306.3130	J-01S	4.7	Manual	T
679104.2450	3083223.2620	J-02S	1.1	Manual	T
679164.8060	3083214.7100	J-06S	0.985	Manual	T
679181.2750	3083178.2880	J-08S	6.1	Manual	T
679213.7730	3083224.9730	J-09S	7.7	Manual	T
679280.5440	3083305.6810	J-10S	2.4	Manual	T
679242.7260	3083326.5280	J-07S	5.1	Manual	T
679225.8560	3083359.9740	J-05S	1.7	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	2.1	Manual	T
679335.0020	3082941.1720	J-21S	4.9	Manual	T
679343.5810	3082969.5980	J-19S	5.25	Manual	T
679394.8070	3082971.8300	J-24S	2.3	Manual	T
679382.8640	3083009.1130	J-20S	5.8	Manual	T
679293.5600	3082950.4980	J-17S	10.6	Manual	T
679360.5700	3083026.4980	J-18S	4.1	Manual	T

679297.0010	3082840.6970	J-23S	9.6	Manual	T
679252.7130	3082781.0290	J-22S	6.85	Manual	T
679222.6340	3082840.1720	J-16S	4.5	Manual	T
679279.6830	3083075.4290	J-14S	4.1	Manual	T
679149.4920	3082933.0980	J-13S	2.4	Manual	T
679272.0040	3082652.6750	J-28S	6.6	Manual	T
679224.5850	3082683.1400	J-26S	2.925	Manual	T
679146.6460	3082549.7640	J-25S	1.6	Manual	T
679169.0760	3082537.3510	J-27S	3.5	Manual	T
679342.7410	3082605.3190	J-35S	9.1	Manual	T
679304.6530	3082548.6880	J-34S	29.3	Manual	T
679382.8900	3082667.5270	J-36S	10.5	Manual	T
679329.4380	3082711.0960	J-29S	12.8	Manual	T
679374.4420	3082791.3300	J-30S	16.6	Manual	T
679453.4760	3082914.1150	J-32S	16.2	Manual	T
679410.1490	3082845.8460	J-31S	17.25	Manual	T
679560.6070	3082897.2580	J-41S	7	Manual	T
679495.8840	3082940.9730	J-33S	4.8	Manual	T
679524.3310	3082886.8990	J-40S	15.8	Manual	T
679497.3310	3082840.3960	J-39S	7.8	Manual	T
679470.3570	3082776.7350	J-38S	16	Manual	T
679433.9450	3082731.6820	J-37S	13.2	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability ($1-\beta$) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

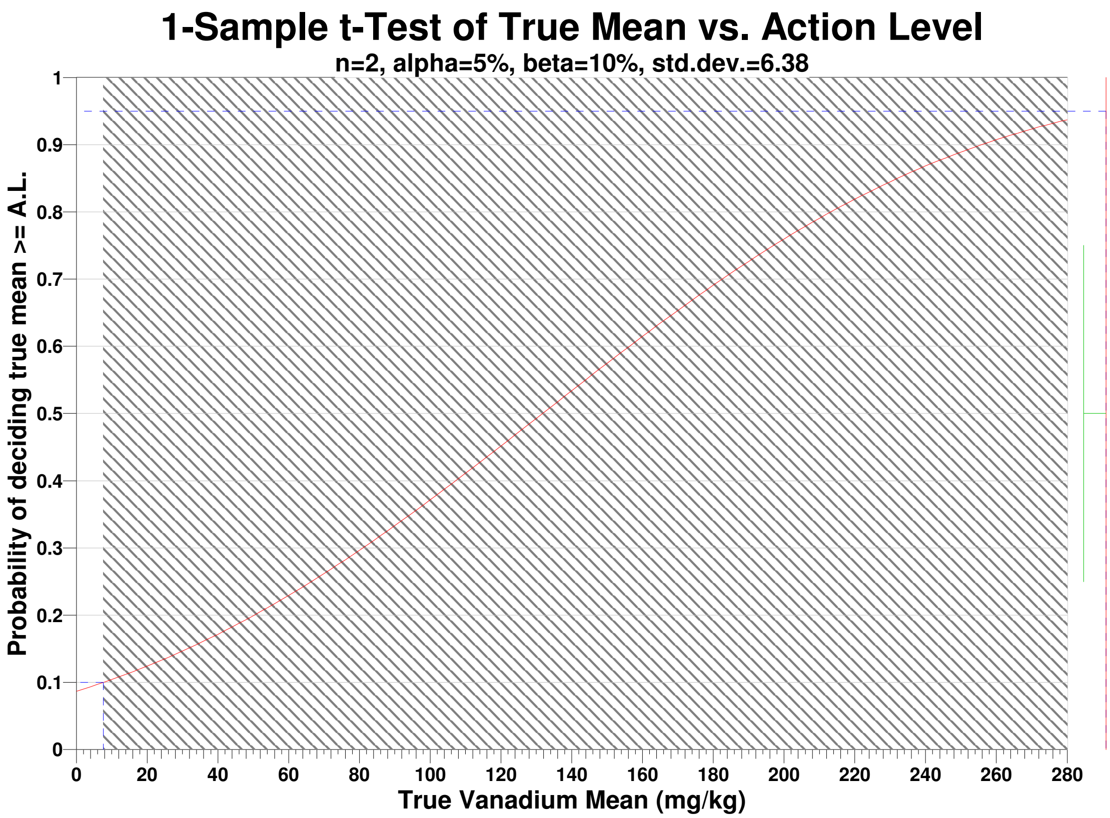
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	Δ	α	β	$Z_{1-\alpha}$ [a]	$Z_{1-\beta}$ [b]
Vanadium	2	6.38 mg/kg	283.38 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at $1-\alpha$. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=291.014		α=5		α=10		α=15	
		s=12.76	s=6.38	s=12.76	s=6.38	s=12.76	s=6.38
LBGR=90	β=5	4	2	3	2	2	1
	β=10	3	2	3	2	2	1
	β=15	3	2	2	2	2	1
LBGR=80	β=5	2	2	2	1	1	1
	β=10	2	2	2	1	1	1
	β=15	2	2	2	1	1	1
LBGR=70	β=5	2	2	2	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Vanadium
The following data points were entered by the user for analysis.

Vanadium (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.985	1.1	1.15	1.3	1.6	1.7	2.1	2.3	2.4	2.4
10	2.925	3.5	4.1	4.1	4.5	4.7	4.8	4.9	5.1	5.1
20	5.25	5.8	6.1	6.6	6.85	7	7.7	7.8	9.1	9.6
30	10.5	10.6	12.8	13.2	15.8	16	16.2	16.6	17.25	22.3
40	29.3									

SUMMARY STATISTICS for Vanadium	
n	41
Min	0.985
Max	29.3
Range	28.315
Mean	7.6368
Median	5.25
Variance	40.68
StdDev	6.3781
Std Error	0.99609
Skewness	1.4681

Interquartile Range				7.8875				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
0.985	1.105	1.36	2.662	5.25	10.55	16.52	21.79	29.3

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Vanadium			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.397	3.05	Yes

The test statistic 3.397 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Vanadium	
1	29.3

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.8816
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

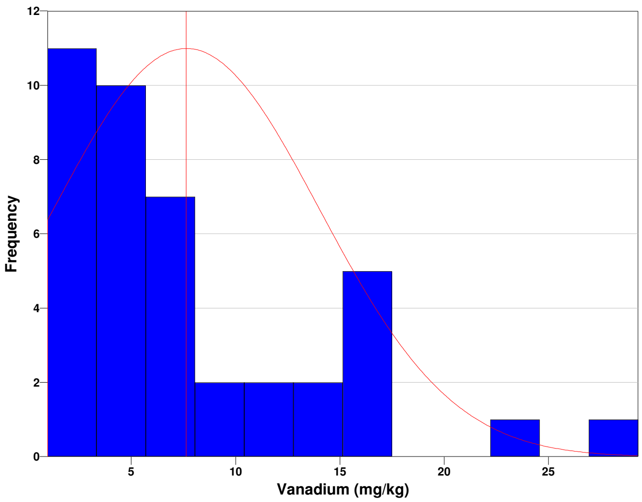
Data Plots for Vanadium
Graphical displays of the data are shown below.

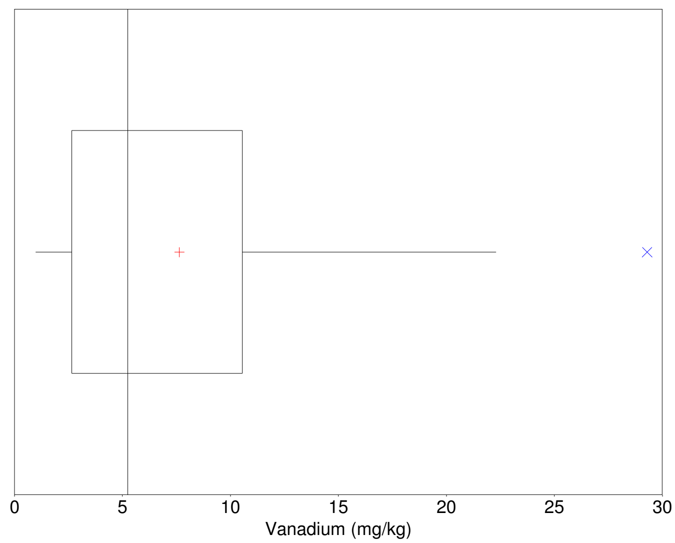
The Histogram is a plot of the fraction of the n observed data that fall within specified data "bins." A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

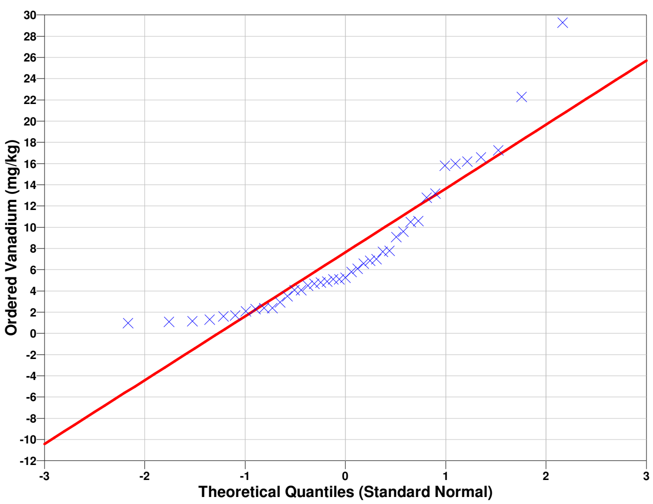
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Vanadium

A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.8546
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	9.314
95% Non-Parametric (Chebyshev) UCL	11.98

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (11.98) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (291.014),
SE is the standard error = (standard deviation) / (square root of n).

This *t* was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-284.49	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.
Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.

Random sampling locations for comparing a mean with a fixed threshold (parametric)

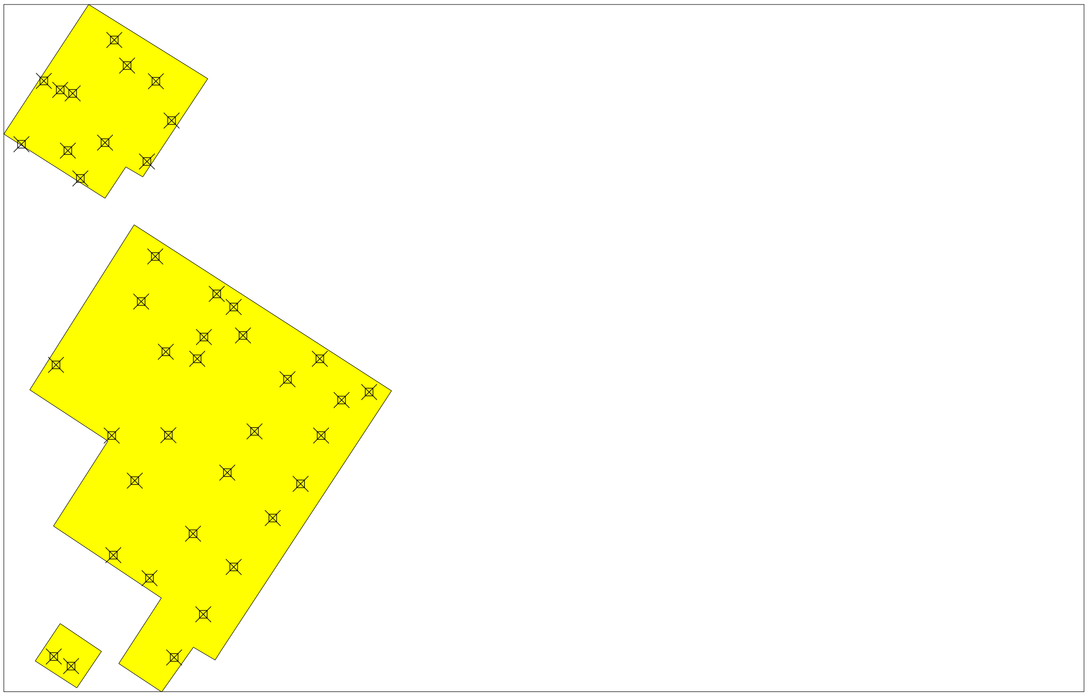
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	25.2	Manual	T
679301.1600	3083254.0340	J-12S	6.1	Manual	T
679171.2970	3083289.7960	J-04S	8.7	Manual	T
679155.0740	3083294.6960	J-03S	11.8	Manual	T
679133.4290	3083306.3130	J-01S	40.7	Manual	T
679104.2450	3083223.2620	J-02S	23.6	Manual	T
679164.8060	3083214.7100	J-06S	3.1	Manual	T
679181.2750	3083178.2880	J-08S	80.3	Manual	T
679213.7730	3083224.9730	J-09S	16.6	Manual	T
679280.5440	3083305.6810	J-10S	22.8	Manual	T
679242.7260	3083326.5280	J-07S	232	Manual	T
679225.8560	3083359.9740	J-05S	10.4	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	31.1	Manual	T
679335.0020	3082941.1720	J-21S	39.5	Manual	T
679343.5810	3082969.5980	J-19S	26.15	Manual	T
679394.8070	3082971.8300	J-24S	22.2	Manual	T
679382.8640	3083009.1130	J-20S	30.2	Manual	T
679293.5600	3082950.4980	J-17S	143	Manual	T
679360.5700	3083026.4980	J-18S	35	Manual	T

679297.0010	3082840.6970	J-23S	44.1	Manual	T
679252.7130	3082781.0290	J-22S	40.65	Manual	T
679222.6340	3082840.1720	J-16S	129	Manual	T
679279.6830	3083075.4290	J-14S	92.6	Manual	T
679149.4920	3082933.0980	J-13S	17.2	Manual	T
679272.0040	3082652.6750	J-28S	48	Manual	T
679224.5850	3082683.1400	J-26S	24.33	Manual	T
679146.6460	3082549.7640	J-25S	59	Manual	T
679169.0760	3082537.3510	J-27S	29.4	Manual	T
679342.7410	3082605.3190	J-35S	20.1	Manual	T
679304.6530	3082548.6880	J-34S	156	Manual	T
679382.8900	3082667.5270	J-36S	40.2	Manual	T
679329.4380	3082711.0960	J-29S	19.7	Manual	T
679374.4420	3082791.3300	J-30S	79	Manual	T
679453.4760	3082914.1150	J-32S	59	Manual	T
679410.1490	3082845.8460	J-31S	26.05	Manual	T
679560.6070	3082897.2580	J-41S	85.3	Manual	T
679495.8840	3082940.9730	J-33S	11.1	Manual	T
679524.3310	3082886.8990	J-40S	31.8	Manual	T
679497.3310	3082840.3960	J-39S	25.9	Manual	T
679470.3570	3082776.7350	J-38S	29.8	Manual	T
679433.9450	3082731.6820	J-37S	48.5	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability (1-β) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n	is the number of samples,
S	is the estimated standard deviation of the measured values including analytical error,
Δ	is the width of the gray region,
α	is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β	is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is 1-α,
$Z_{1-\beta}$	is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is 1-β.

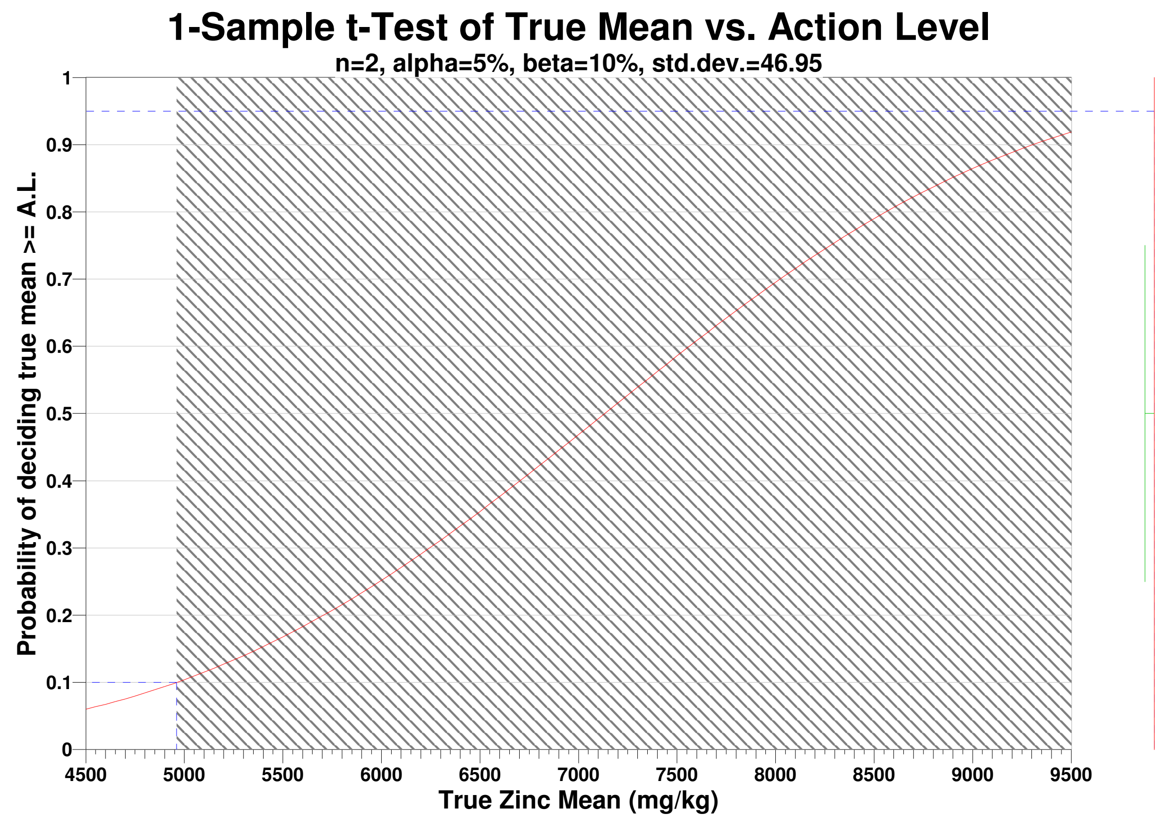
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Zinc	2	46.95 mg/kg	4960.74 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at 1-α on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at 1-α. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions
The assumptions associated with the formulas for computing the number of samples are:
1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis
The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=9921.47		α=5		α=10		α=15	
		s=93.9	s=46.95	s=93.9	s=46.95	s=93.9	s=46.95
LBGR=90	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=80	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	β=5	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Zinc
The following data points were entered by the user for analysis.

Zinc (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	3.1	6.1	8.7	10.4	11.1	11.8	16.6	17.2	19.7	20.1
10	22.2	22.8	23.6	24.33	25.2	25.9	26.05	26.15	29.4	29.8
20	30.2	31.1	31.8	35	39.5	40.2	40.65	40.7	44.1	48
30	48.5	59	59	79	80.3	85.3	92.6	129	143	156
40	232									

SUMMARY STATISTICS for Zinc	
n	41
Min	3.1
Max	232
Range	228.9
Mean	46.956
Median	30.2
Variance	2149.9
StdDev	46.367
Std Error	7.2413
Skewness	2.3039

Interquartile Range				32.6				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
3.1	6.36	10.54	21.15	30.2	53.75	121.7	154.7	232

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Zinc			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.991	3.05	Yes

The test statistic 3.991 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Zinc	
1	232

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.7959
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

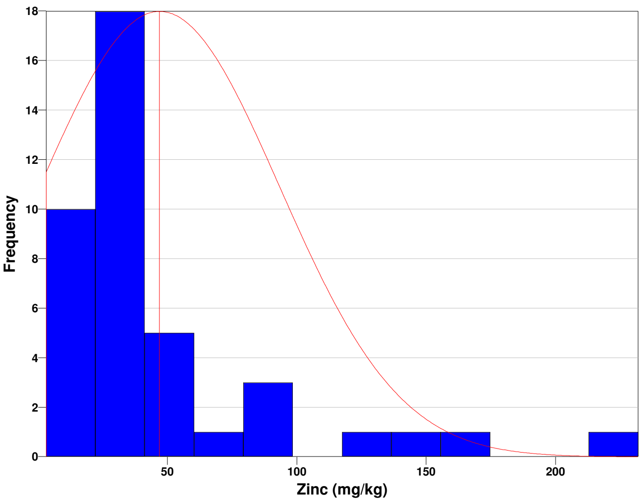
Data Plots for Zinc
Graphical displays of the data are shown below.

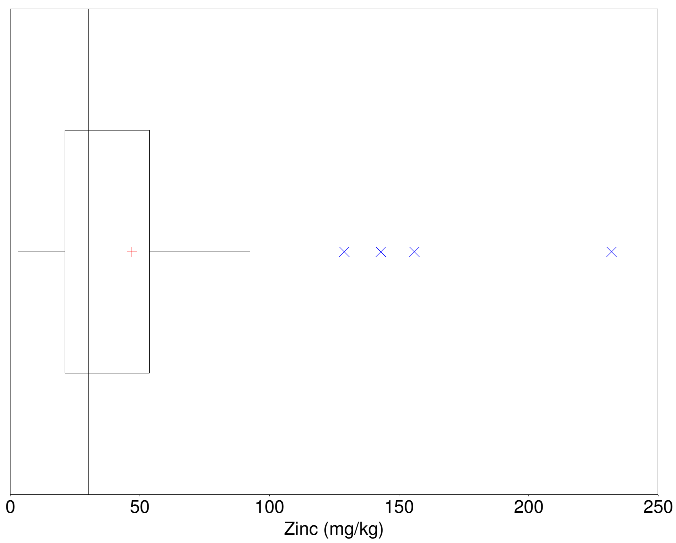
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

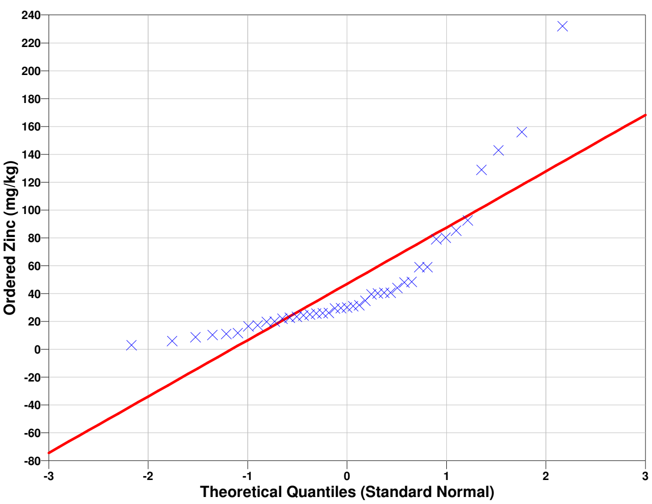
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25th and 75th percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Zinc
A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.7426
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean
Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	59.15
95% Non-Parametric (Chebyshev) UCL	78.52

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (78.52) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test
A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value t was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where
x is the sample mean of the n=41 data,
AL is the action level or threshold (9921.47),
SE is the standard error = (standard deviation) / (square root of n).

This t was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-1363.6	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

Random sampling locations for comparing a mean with a fixed threshold (parametric)

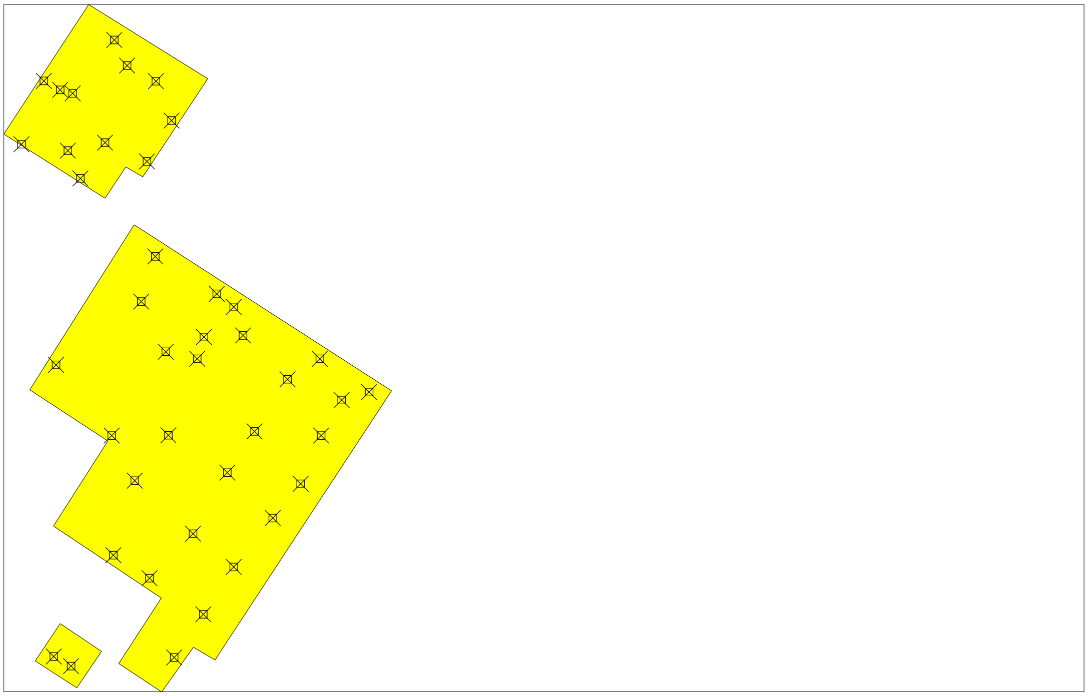
Summary

This report summarizes the sampling design, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	2
Number of samples on map [a]	41
Number of selected sample areas [b]	2
Specified sampling area [c]	188054.34 m^2
Total cost of sampling [d]	$2,000.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
679268.7700	3083200.3260	J-11S	25.2	Manual	T
679301.1600	3083254.0340	J-12S	6.1	Manual	T
679171.2970	3083289.7960	J-04S	8.7	Manual	T
679155.0740	3083294.6960	J-03S	11.8	Manual	T
679133.4290	3083306.3130	J-01S	40.7	Manual	T
679104.2450	3083223.2620	J-02S	23.6	Manual	T
679164.8060	3083214.7100	J-06S	3.1	Manual	T
679181.2750	3083178.2880	J-08S	80.3	Manual	T
679213.7730	3083224.9730	J-09S	16.6	Manual	T
679280.5440	3083305.6810	J-10S	22.8	Manual	T
679242.7260	3083326.5280	J-07S	232	Manual	T
679225.8560	3083359.9740	J-05S	10.4	Manual	T

Area: Area 3					
X Coord	Y Coord	Label	Value	Type	Historical
679261.0980	3083016.3510	J-15S	31.1	Manual	T
679335.0020	3082941.1720	J-21S	39.5	Manual	T
679343.5810	3082969.5980	J-19S	26.15	Manual	T
679394.8070	3082971.8300	J-24S	22.2	Manual	T
679382.8640	3083009.1130	J-20S	30.2	Manual	T
679293.5600	3082950.4980	J-17S	143	Manual	T
679360.5700	3083026.4980	J-18S	35	Manual	T

679297.0010	3082840.6970	J-23S	44.1	Manual	T
679252.7130	3082781.0290	J-22S	40.65	Manual	T
679222.6340	3082840.1720	J-16S	129	Manual	T
679279.6830	3083075.4290	J-14S	92.6	Manual	T
679149.4920	3082933.0980	J-13S	17.2	Manual	T
679272.0040	3082652.6750	J-28S	48	Manual	T
679224.5850	3082683.1400	J-26S	24.33	Manual	T
679146.6460	3082549.7640	J-25S	59	Manual	T
679169.0760	3082537.3510	J-27S	29.4	Manual	T
679342.7410	3082605.3190	J-35S	20.1	Manual	T
679304.6530	3082548.6880	J-34S	156	Manual	T
679382.8900	3082667.5270	J-36S	40.2	Manual	T
679329.4380	3082711.0960	J-29S	19.7	Manual	T
679374.4420	3082791.3300	J-30S	79	Manual	T
679453.4760	3082914.1150	J-32S	59	Manual	T
679410.1490	3082845.8460	J-31S	26.05	Manual	T
679560.6070	3082897.2580	J-41S	85.3	Manual	T
679495.8840	3082940.9730	J-33S	11.1	Manual	T
679524.3310	3082886.8990	J-40S	31.8	Manual	T
679497.3310	3082840.3960	J-39S	25.9	Manual	T
679470.3570	3082776.7350	J-38S	29.8	Manual	T
679433.9450	3082731.6820	J-37S	48.5	Manual	T

Primary Sampling Objective
The primary purpose of sampling at this site is to compare a mean value of a site with a fixed threshold. The working hypothesis (or 'null' hypothesis) is that the mean value at the site is equal to or exceeds the threshold. The alternative hypothesis is that the mean value is less than the threshold. VSP calculates the number of samples required to reject the null hypothesis in favor of the alternative hypothesis, given a selected sampling approach and inputs to the associated equation.

Selected Sampling Approach
A parametric random sampling approach was used to determine the number of samples and to specify sampling locations. A parametric formula was chosen because the conceptual model and historical information (e.g., historical data from this site or a very similar site) indicate that parametric assumptions are reasonable. These assumptions will be examined in post-sampling data analysis.

Both parametric and non-parametric approaches rely on assumptions about the population. However, non-parametric approaches typically require fewer assumptions and allow for more uncertainty about the statistical distribution of values at the site. The trade-off is that if the parametric assumptions are valid, the required number of samples is usually less than the number of samples required by non-parametric approaches.

Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Number of Total Samples: Calculation Equation and Inputs
The equation used to calculate the number of samples is based on a Student's t-test. For this site, the null hypothesis is rejected in favor of the alternative hypothesis if the sample mean is sufficiently smaller than the threshold. The number of

samples to collect is calculated so that 1) there will be a high probability (1-β) of rejecting the null hypothesis if the alternative hypothesis is true and 2) a low probability (α) of rejecting the null hypothesis if the null hypothesis is true.

The formula used to calculate the number of samples is:

$$n = \frac{S^2}{\Delta^2}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 + 0.5Z_{1-\alpha}^2$$

where

n is the number of samples,
S is the estimated standard deviation of the measured values including analytical error,
Δ is the width of the gray region,
α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,
$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is 1-α,
$Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is 1-β.

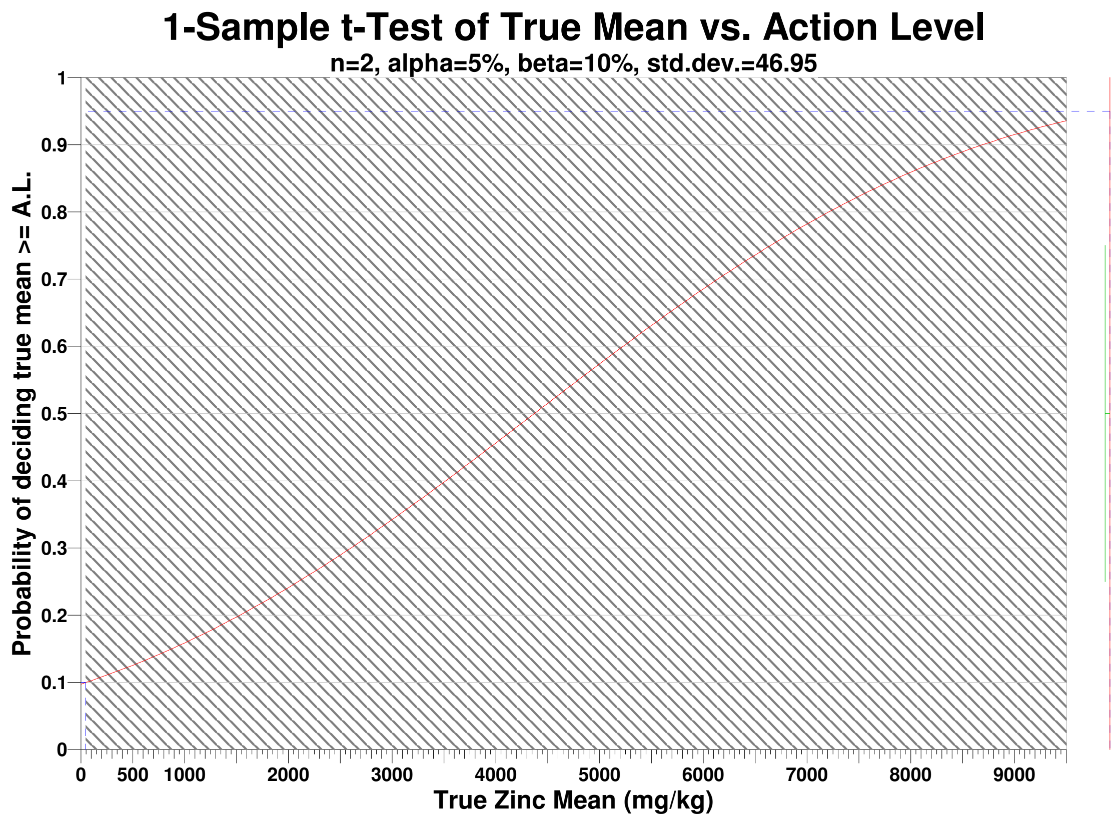
The values of these inputs that result in the calculated number of sampling locations are:

Analyte	n	Parameter					
		S	**Δ**	**α**	**β**	**$Z_{1-\alpha}$** [a]	**$Z_{1-\beta}$** [b]
Zinc	2	46.95 mg/kg	9872.91 mg/kg	0.05	0.1	1.64485	1.28155

[a] This value is automatically calculated by VSP based upon the user defined value of α.
[b] This value is automatically calculated by VSP based upon the user defined value of β.

The following figure is a performance goal diagram, described in EPA's QA/G-4 guidance (EPA, 2000). It shows the probability of concluding the sample area is dirty on the vertical axis versus a range of possible true mean values for the site on the horizontal axis. This graph contains all of the inputs to the number of samples equation and pictorially represents the calculation.

The red vertical line is shown at the threshold (action limit) on the horizontal axis. The width of the gray shaded area is equal to Δ; the upper horizontal dashed blue line is positioned at 1-α on the vertical axis; the lower horizontal dashed blue line is positioned at β on the vertical axis. The vertical green line is positioned at one standard deviation below the threshold. The shape of the red curve corresponds to the estimates of variability. The calculated number of samples results in the curve that passes through the lower bound of Δ at β and the upper bound of Δ at 1-α. If any of the inputs change, the number of samples that result in the correct curve changes.



Statistical Assumptions

The assumptions associated with the formulas for computing the number of samples are:

1. the sample mean is normally distributed (this happens if the data are roughly symmetric or the sample size is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed),
2. the variance estimate, S^2, is reasonable and representative of the population being sampled,
3. the population values are not spatially or temporally correlated, and
4. the sampling locations will be selected randomly.

The first three assumptions will be assessed in a post data collection analysis. The last assumption is valid because the sample locations were selected using a random process.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying the standard deviation, lower bound of gray region (% of action level), beta (%), probability of mistakenly concluding that μ > action level and alpha (%), probability of mistakenly concluding that μ < action level. The following table shows the results of this analysis.

Number of Samples							
AL=9921.47		α=5		α=10		α=15	
		s=93.9	s=46.95	s=93.9	s=46.95	s=93.9	s=46.95
LBGR=90	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=80	β=5	2	2	1	1	1	1
	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1
LBGR=70	β=5	2	2	1	1	1	1

	β=10	2	2	1	1	1	1
	β=15	2	2	1	1	1	1

s = Standard Deviation
LBGR = Lower Bound of Gray Region (% of Action Level)
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $2,000.00, which averages out to a per sample cost of $1,000.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**2 Samples**
Field collection costs		$100.00	$200.00
Analytical costs	$400.00	$400.00	$800.00
Sum of Field & Analytical costs		**$500.00**	**$1,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$2,000.00**

Data Analysis for Zinc
The following data points were entered by the user for analysis.

Zinc (mg/kg)										
Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	3.1	6.1	8.7	10.4	11.1	11.8	16.6	17.2	19.7	20.1
10	22.2	22.8	23.6	24.33	25.2	25.9	26.05	26.15	29.4	29.8
20	30.2	31.1	31.8	35	39.5	40.2	40.65	40.7	44.1	48
30	48.5	59	59	79	80.3	85.3	92.6	129	143	156
40	232									

SUMMARY STATISTICS for Zinc	
n	41
Min	3.1
Max	232
Range	228.9
Mean	46.956
Median	30.2
Variance	2149.9
StdDev	46.367
Std Error	7.2413
Skewness	2.3039

Interquartile Range				32.6				
Percentiles								
1%	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**	**99%**
3.1	6.36	10.54	21.15	30.2	53.75	121.7	154.7	232

Outlier Test
Rosner's test for multiple outliers was performed to test whether the most extreme value is a statistical outlier. The test was conducted at the 5% significance level.

Data should not be excluded from analysis solely on the basis of the results of this or any other statistical test. If any values are flagged as possible outliers, further investigation is recommended to determine whether there is a plausible explanation that justifies removing or replacing them.

In using Rosner's test to detect up to 1 outlier, a test statistic R_1 is calculated, and compared with a critical value C_1 to test the hypothesis that there is one outlier in the data.

ROSNER'S OUTLIER TEST for Zinc			
k	**Test Statistic R_k**	**5% Critical Value C_k**	**Significant?**
1	3.991	3.05	Yes

The test statistic 3.991 exceeded the corresponding critical value, therefore that test is significant and we conclude that the most extreme value is an outlier at the 5% significance level.

SUSPECTED OUTLIERS for Zinc	
1	232

A normal distribution test indicated that the data do not appear to be normally distributed, so further investigation is recommended before using the results of this test. Because Rosner's test can be used only when the data without the suspected outlier are approximately normally distributed, a Shapiro-Wilk test for normality was performed at a 5% significance level.

NORMAL DISTRIBUTION TEST (excluding outliers)	
Shapiro-Wilk Test Statistic	0.7959
Shapiro-Wilk 5% Critical Value	0.94

The calculated Shapiro-Wilk test statistic is less than the 5% Shapiro-Wilk critical value, so the test rejects the hypothesis that the data are normal and concludes that the data, excluding the most extreme value, do not appear to follow a normal distribution at the 5% level of significance. Rosner's test may not be appropriate if the assumption of normally distributed data is not justified for this data set. Examine the Q-Q plot displayed below to further assess the normality of the data.

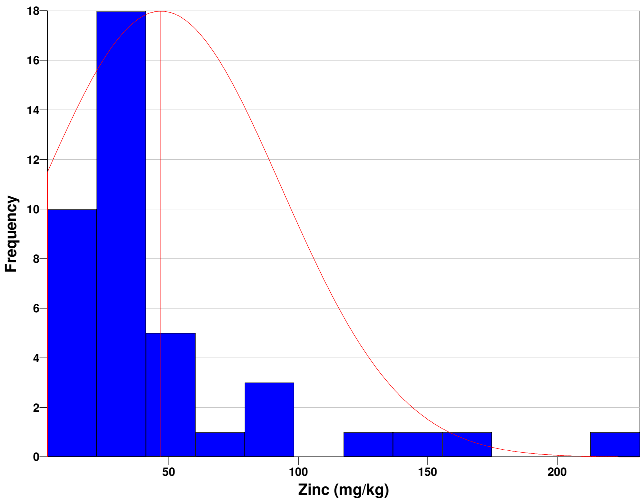
Data Plots for Zinc
Graphical displays of the data are shown below.

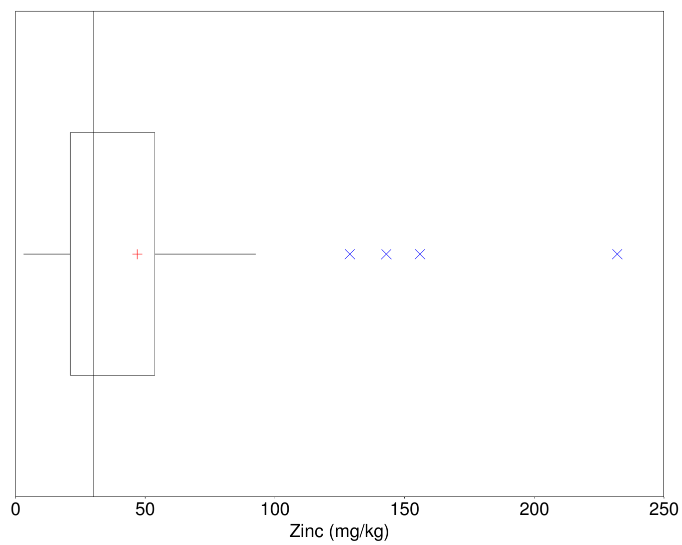
The Histogram is a plot of the fraction of the n observed data that fall within specified data “bins.” A histogram is generated by dividing the x axis (range of the observed data values) into "bins" and displaying the number of data in each bin as the height of a bar for the bin. The area of the bar is the fraction of the n data values that lie within the bin. The sum of the fractions for all bins equals one. A histogram is used to assess how the n data are distributed (spread) over their range of values. If the histogram is more or less symmetric and bell shaped, then the data may be normally distributed.

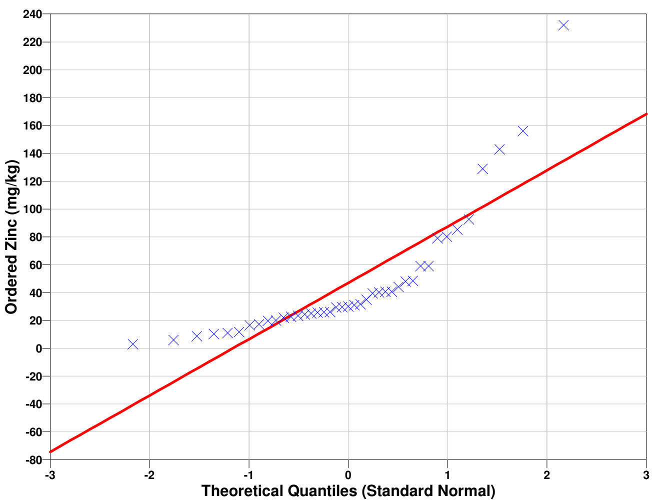
The Box and Whiskers plot is composed of a central box divided by a line, and with two lines extending out from the box, called the "whiskers". The line through the box is drawn at the median of the n data observed. The two ends of the box represent the 25^{th} and 75^{th} percentiles of the n data values, which are also called the lower and upper quartiles, respectively, of the data set. The sample mean (mean of the n data) is shown as a "+" sign. The upper whisker extends

to the largest data value that is less than the upper quartile plus 1.5 times the interquartile range (upper quartile minus the lower quartile). The lower whisker extends to the smallest data value that is greater than the lower quartile minus 1.5 times the interquartile range. Extreme data values (greater or smaller than the ends of the whiskers) are plotted individually as blue Xs. A Box and Whiskers plot is used to assess the symmetry of the distribution of the data set. If the distribution is symmetrical, the box is divided into two equal halves by the median, the whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

The Q-Q plot graphs the quantiles of a set of n data against the quantiles of a specific distribution. We show here only the Q-Q plot for an assumed normal distribution. The p^{th} quantile of a distribution of data is the data value, x_p, for which a fraction p of the distribution is less than x_p. If the data plotted on the normal distribution Q-Q plot closely follow a straight line, even at the ends of the line, then the data may be assumed to be normally distributed. If the data points deviate substantially from a linear line, then the data are not normally distributed.







For more information on these plots consult Guidance for Data Quality Assessment, EPA QA/G-9, pgs 2.3-1 through 2.3-12. (http://www.epa.gov/quality/qa-docs.html).

Tests for Zinc

A goodness-of-fit test was performed to test whether the data set had been drawn from an underlying normal distribution. The Shapiro-Wilk (SW) test was used to test the null hypothesis that the data are normally distributed. The test was conducted at the 5% significance level, i.e., the probability the test incorrectly rejects the null hypothesis was set at 0.05.

NORMAL DISTRIBUTION TEST	
Shapiro-Wilk Test Statistic	0.7426
Shapiro-Wilk 5% Critical Value	0.941

The calculated SW test statistic is less than the 5% Shapiro-Wilk critical value, so we can reject the hypothesis that the data are normal, or in other words the data do not appear to follow a normal distribution at the 5% level of significance. The Q-Q plot displayed above should be used to further assess the normality of the data.

Upper Confidence Limit on the True Mean

Two methods were used to compute the upper confidence limit (UCL) on the mean. The first is a parametric method that assumes a normal distribution. The second is the Chebyshev method, which requires no distributional assumption.

UCLs ON THE MEAN	
95% Parametric UCL	59.15
95% Non-Parametric (Chebyshev) UCL	78.52

Because the data do not appear to be normally distributed according to the goodness-of-fit test performed above, the non-parametric UCL (78.52) may be a more accurate upper confidence limit on the true mean.

One-Sample t-Test

A one-sample t-test was performed to compare the sample mean to the action level. The null hypothesis used is that the true mean equals or exceeds the action level (AL). The t-test was conducted at the 5% significance level. The sample value *t* was computed using the following equation:

$$t = \frac{\bar{x} - AL}{SE}$$

where

x is the sample mean of the n=41 data,

AL is the action level or threshold (9921.47),

SE is the standard error = (standard deviation) / (square root of n).

This *t* was then compared with the critical value $t_{0.95}$, where $t_{0.95}$ is the value of the t distribution with n-1=40 degrees of freedom for which the proportion of the distribution to the left of $t_{0.95}$ is 0.95. The null hypothesis will be rejected if $t < -t_{0.95}$.

ONE-SAMPLE t-TEST		
t-statistic	**Critical Value $t_{0.95}$**	**Null Hypothesis**
-1363.6	1.6839	Reject

The test rejected the null hypothesis that the mean value at the site exceeds the threshold, therefore conclude the true mean is less than the threshold.

Because the data do not appear to be normally distributed, the MARSSIM Sign Test might be preferred over the One Sample t-Test. The following table represents the results of the MARSSIM Sign Test using the current data:

MARSSIM Sign Test		
Test Statistic (S+)	**95% Critical Value**	**Null Hypothesis**
41	26	Reject

This report was automatically produced* by Visual Sample Plan (VSP) software version 5.3.1.

Software and documentation available at http://dqo.pnl.gov/vsp

* - The report contents may have been modified or reformatted by end-user of software.